PEOPLE PERFORMANCE

Dependability and Safety Instrument (DSI) Version 1.1

Technical Manual

Dependability and Safety Instrument (DSI) Version 1.1

› Technical Manual

Helping organisations improve customer service and reduce workplace accidents

Eugene Burke, Carly Vaughan
and Hannah Ablitt

© 2010, SHL Group Limited
www.shl.com

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means or stored in a database or retrieval system without the prior written permission of SHL Group Limited.

Contents

What you will find in this manual	**6**
Acknowledgments	**8**
Dependability and outcomes in the workplace	**9**
The cost of absenteeism and poor employee attendance	9
The cost of accidents in the workplace	9
The cost of delivering poor customer service	9
Different business issues but a common set of underlying behavioural causes	10
Defining dependability and associated workplace behaviours	**11**
The role of personality in OCBs and CWBs	**15**
General relationships between personality and OCBs and CWBs	15
Personality and service (customer) orientation	15
Personality and accidents	16
Digman's higher order factor Alpha	16
Towards a systemic model for predicting workplace outcomes: Linking disposition, dependability, customer service and accident proneness	**19**
Using the model to predict customer service outcomes	20
Using the model to predict safety outcomes	20
Using the model to predict overall perceived value of employees	20
The evidence presented in the next two sections of this manual	21
Dependability and manager and supervisor perceptions of employees	**22**
Predicting outcomes in customer service roles	22
Predicting outcomes in safety critical roles	23
A summary of the relationships between dependability and workplace outcomes	24

The construction of DSI and evidence supporting its criterion validity **26**

The construction and scoring of DSI 26

Revision of DSI and Version 1.1 27

A meta-analysis of DSI criterion validity 28

The case of unauthorised absence and customer care service advisers in the energy industry 32

The case of security guards, absenteeism, accidents and incidents of attacks 32

Understanding why DSI works: Evidence of construct validity for DSI scores **34**

Automotive engineers and the relationship between DSI scores and WSQ scales 34

OPQ32 and the relationship between DSI scores and Big 5 indicators 35

International bank call centre and the relationship between DSI and the Customer Contact Styles Questionnaire (CCSQ) 38

Relationship between DSI and cognitive ability test scores 39

Setting DSI score bands to provide levels of risk management in screening potential employees **40**

Reliability and fairness of DSI scores **42**

The reliability of DSI scores 42

Evaluating the fairness of DSI scores 43

Evaluating differential item functioning (DIF) of DSI items for English fluency 44

Evaluating differential item functioning (DIF) of DSI items for Demographic Groups 46

Evaluating adverse (disparate) impact of applying DSI risk bands 48

Age and DSI scores 52

A summary of findings on bias and adverse impact analyses of DSI 53

Faking and DSI **54**

Using DSI as a human factors audit to provide data on risks in organisations **56**

References **57**

What you will find in this manual

This manual provides evidence gathered through the scientific programme supporting the Dependability and Safety Instrument (DSI). This manual replaces the previous 2006 manual for DSI Version 1.0 and reflects the upgrade to DSI Version 1.1 following a series of large scale analyses, learning since the launch of Version 1.0 and discussions with clients in Asia, Europe, North America and South Africa. This manual covers the English language version of the DSI and further supplements will be issued as other language versions become available through the localisation programme now in place.

The DSI scientific programme has shown that:

- Behaviours that define dependability in the workplace are important for good attendance, customer service and safety at work, and play a key role in the judgements made by supervisors and managers about who represents an effective employee and who does not.
- The SHL definition of dependability generalises across different organisations and industry sectors, public or private sector organisations, as well as different countries, and consistently relates to outcomes in the workplace whether that is in customer facing roles or safety critical roles.
- These behaviours are consistently predicted by the score from a forced choice questionnaire, DSI, which originally comprised 22 statement pairs (Version 1.0) but has been made more efficient with 18 statement pairs (Version 1.1) after an extensive review programme.
- The original classification of scores into three bands of risk (red, amber and green) can be refined to five levels of risk to allow clients greater flexibility in the setting of cut-scores when used in the recruitment and selection of personnel. The five risk levels or score bands are another feature of Version 1.1 resulting from research since 2006.
- DSI scores are stable over time (as measured using a test-retest or stability coefficient) and the tool meets most definitions of fair assessments such as the 80% or 4/5th's rule in the United States in showing no adverse impact against women, older candidates or candidates from ethnic minorities. Furthermore, analysis has shown that the questionnaire is suitable for use with those with lower levels of educational attainment and with a reasonable fluency with English (as mentioned, a localisation programme is in place to provide DSI in other languages and these will be covered by future supplements to this manual and as other language versions of DSI are released).
- The DSI can be deployed via paper-and-pencil, telephone and online administration with no reduction in the quality of the assessment.

The information that the reader will find in this manual covers the business issues that DSI was designed to address, the general research literature that guided DSI's development as well as the evidence gathered through SHL's research. The evidence that reader will find in this manual includes:

- Validation of the behavioural criterion measures used to operationalise SHL's definition of dependable behaviours.
- Criterion validation of DSI against the behavioural measures of dependability including a meta-analysis evaluating the generalisability of DSI validities across organisations, roles (jobs) and geographies (countries).
- Construct validations that help to explain why DSI offers consistent predictions of dependable workplace behaviours.
- Case studies showing the relationship between DSI scores and indicators of counter-productive work behaviours (CWB) such as unauthorised absence and accidents for which the employee was responsible.
- Reliability analysis showing the stability of scores over time, as well as comparisons of DSI scores obtained from high and low stakes conditions to evaluate how robust DSI is to manipulation by candidates.
- Analyses of DSI scores by the demographics of gender, age, ethnicity and educational attainment to evaluate adverse impact and fairness of cut-scores.
- Bias analysis that explores the performance of DSI items for any differences in functioning by gender, age, ethnicity and language fluency.

The scientific programme supporting DSI has, to date, evaluated data on over 6,000 people across four countries and multiple organisations. Data supporting the criterion validity of DSI thus far amounts to 898 employees across 13 organisations covering customer facing and safety critical roles in Australia, North America, the UK and the US. We are committed to an ongoing programme of data collection and evaluation of DSI, so please contact us if you would like to participate in this programme. Contact details of your local SHL office are available from www.shl.com.

Acknowledgments

A number of people and organisations have been involved in the development and trialling of the Dependability and Safety Instrument. We would like to express our thanks to them for their support and assistance, and hope that this manual does justice to their investment in the development of DSI.

We would especially like to thank Lesley Kirby who was co-author of Version 1.0 of DSI, and to Paul Levett and Simon Raymond for sponsoring the original DSI development programme that led to Version 1.0. The new Version of DSI owes much to the work of Claire Fix who managed and executed the programme that led to Version 1.1. We would also like to thank Kim Dowdeswell, Tim Irvine and Nadene Venter for their energies in realising Version 1.1.

There are many organisations in Australia, the UK and the US that have contributed to the development of the DSI since the programme was first initiated in 2004. We would like to thank all of these organisations for their contribution to the DSI programme.

Dependability and outcomes in the workplace

We have defined dependability as a set of behaviours related to time keeping, meeting expectations for how to behave in the workplace (e.g. compliance with procedures and organisational policies), getting along with and supporting work colleagues, and coping with the day-to-day challenges that normally occur in the workplace. We will provide more information on our definition of dependability in the next section of this manual. In this section, we will briefly explore the organisational impacts of dependability or rather its shadow side in terms of unreliable or irresponsible behaviours manifested in the workplace.

The cost of absenteeism and poor employee attendance

Slora (1991) conducted a series of surveys with fast food and supermarket employees to explore the extent to which those employees admitted to counter-productive behaviours. The results showed that 96% of fast food workers and 94% of supermarket workers admitted to some form of counter-productive behaviours, with lateness (71% and 70%) and arguing with supervisors (78% and 61%) the most commonly reported of these behaviours.

The UK Confederation of British Industry (CBI) estimated that in 2004, the UK economy lost £11.6 ($16.2) million due to unauthorised absence from work. In a subsequent publication in 2007, the CBI's Absence Report estimated that the cost of absence from work had increased to £13.2 ($18.5) billion.

The cost of accidents in the workplace

The UK Health and Safety Executive (HSE) reported in 2004 that workplace accidents and work-related ill health cost employers between £3.9 ($5.5) and £7.8 ($10.9) billion in 2001 and 2002. Clarke and Robertson (2008) cite HSE statistics for 2003 and 2004 showing that the UK economy lost 39 million working days due to accidents, of which 9 million days were due to workplace injuries. Clarke and Robertson also cite an estimated cost of workplace accidents to the US economy of $156 (£111.4) billion in 2003. In short, while statistics and costs vary, various sources are consistent in suggesting that accidents in the workplace remain a significant issue and one that results in substantial financial and human costs.

In 2004, the Future Foundation (2004) conducted an international survey of 2,500 workers in 7 countries and discovered that over 70% of mistakes made by employees in the workplace were hidden from supervisors and managers, suggesting a significant blind spot in all organisations.

The cost of delivering poor customer service

Goodman (1999) reports that the Technical Assistance Research Program (TARP) found that, on average, customer loyalty drops by 20% if the customer has encountered a problem with a product or a service. They also found that people tend to pay more attention to bad word of mouth such that twice as many people hear about a bad customer experience as they do about a good experience.

In November 2007, a UK YouGov (see Confederation of British Industry, 2007b) poll revealed that 48% of British adults believe that excellent customer service is the most important characteristic for a company's reputation, and that 58% of consumers are willing to pay more for the same product when purchased from their most highly regarded company.

Different business issues but a common set of underlying behavioural causes

It might surprise the reader that there appears to be a common set of employee behaviours that influence a range of outcomes in the workplace such as whether the employee will generally have a good attendance record, or be effective in a customer service role, or operate effectively where safety is important.

This view of correlated behaviours underpinning different outcomes is supported by the growing research literature on counter-productive work behaviours and organisational citizenship behaviours. For example, Viswesvaran (2002) cites results that are consistent with the view that higher absenteeism, for example, could be an indicator of an employee or organisational member withdrawing effort from work tasks. This view of correlated behaviours underpinning multiple outcomes in the workplace has been referred to as the co-occurrence of counter-productive behaviours by those such as Gruys and Sackett (2003). This means that the manifestation of one behaviour is a potential indicator that other behaviours are also more likely to be exhibited. We have also found evidence of correlated behaviours underlying different outcomes across different organisations, and we have based our definition of dependability on these behaviours. DSI was designed to predict these behaviours and we will now describe them in more detail in the next section of this manual.

Defining dependability and associated workplace behaviours

Our knowledge of what underpins effective performance and counter-productive behaviours in the workplace has grown significantly in the past decade. A more traditional view focused attention and research on task performance (i.e. how quickly employees acquired and demonstrated the skills stated in a job description). More recent research has broadened the view of effective performance to include contextual behaviours that influence levels of effort and commitment that employees will invest in an organisation or work team.

This research can be broadly broken into two strands. Organisational Citizenship Behaviour (OCB) research has explored behaviours that contribute to the social functioning of organisations (e.g. Borman and Motowidlo, 1997). They have been variously labelled as prosocial organisational behaviour and extra role behaviour, and have been generally characterised as behaviours that evidence employees "going the extra mile" in support of the organisation and its goals.

Examples of OCBs include altruism, civic virtue, courtesy, sportsmanship (not complaining about small or trivial matters) and conscientiousness, though the latter relates in OCB terms to compliance with organisational expectations and norms rather than, as in Big 5 definitions, quality, structure, being organised and achievement orientation (LePine, Erez and Johnson, 2002).

A parallel line of research has focused on Counter-productive Work Behaviours (CWBs). Sackett and DeVore (2001) have defined CWBs as "... any intentional behaviour on the part of an organizational member viewed by the organization as contrary to its legitimate interest". Researchers in industrial sociology and organisational behaviour such as Ackroyd and Thompson (2003) have explored the symptoms and antecedents of what they describe as "... employees doing what they are not supposed to do".

Examples of CWBs include appropriation of time (e.g. time wasting and absenteeism), appropriation of work or effort (an extreme being sabotage but milder symptoms being manipulating how effort is recorded and rewarded), appropriation of product (extremes being theft and pilferage) and appropriation of identity (such as creating a work group identity that conflicts with the goals and identity of the organisation).

Research shows that, while CWBs and OCBs tend to look at different aspects of workplace behaviour, they are generally correlated and in the negative direction that one would expect (Berry, Ones and Sackett, 2007; Gruys, 1999; Sackett, 2002).

These respective lines of research framed our development programme and the first step of that programme was to construct a series of criterion scales that provided concrete definitions of dependability, and in such a way as to define OCB and CWB aspects for each set of behaviours. Our initial work was stimulated by the taxonomy proposed by Ackroyd and Thompson (2003) which covers aspects of the contract between organisation and employee for use of time, use of resources and relationships between the employee and other employees, and between the employee and the employer.

DSI validation studies conducted since 2004 provided data on 898 employees in various organisational settings and roles. An exploratory maximum likelihood factor analysis[1] of these manager ratings showed a four factor oblique model to offer an adequate fit to data on 10 Likert style items (Kaiser-Meyer-Olkin, KMO, of 0.841 and chi-square goodness of fit of 0.083). These factors are summarised in Table 1 below with descriptions of the OCB and CWB aspects of each behavioural item.

Table 1: The two faces of four dependable workplace behaviours

Aspect	Cluster	Behaviours		
OCB	Time Keeping	Rarely has time off	Arrives for work on time	Returns from breaks on time
CWB	Time Keeping	Frequently has time off	Frequently late for work	Often returns from breaks late
OCB	Meeting Expectations	Sticks to company regulations	Checks their work for mistakes	
CWB	Meeting Expectations	Does not stick to company regulations	Does not check their work for mistakes	
OCB	Working with Others	Rarely has disagreements with colleagues	Keeps an even temper in most situations	
CWB	Working with Others	Often has disagreements with colleagues	Rarely keeps an even temper	
OCB	Coping with Pressure	Is confident about their own abilities	Handles stressful situations well	Can handle situations of conflict well
CWB	Coping with Pressure	Lacks confidence in their own abilities	Does not handle stressful situations well	Does not handle situations of conflict well

The four clusters shown in Table 1 differ slightly to those originally reported by Burke and Kirby (2006) in the technical manual for Version 1.0 of DSI. 'Coping with Pressure' was originally titled 'Being Confident and Delivering', and this change in title reflects the clarification of scales shown in Table 1. 'Time Keeping' retains the two items that were originally assigned to a cluster labelled 'Being Reliable', but has an additional behaviour, 'returns from breaks on time', re-assigned from the original Version 1.0

[1]The aim of the analysis used here was to identify the minimum number of factors required to explain the covariances between items.

scale 'Complying with Policies and Procedures'. The latter scale has been re-titled 'Meeting Expectations' and now comprises the two items shown in Table 1.

The changes in scale compositions largely reflect the larger sample of 898 now available in contrast to the original 2006 sample of 221. The more recent sample also includes a wider range of jobs and roles, organisational and workplace settings as well as nationalities and geographies, and is therefore more likely to offer a valid representation of the structure of managers' perceptions of effective and ineffective behaviours in operational roles.

Table 2: Intercorrelations between dependability clusters

Cluster	Time Keeping (TK)	Meeting Expectations (ME)	Working with Others (WWO)	Coping With Pressure (CWP)
TK	***0.74***	**0.70**	**0.32**	**0.33**
ME		***0.75***	**0.41**	**0.55**
WWO			***0.76***	**0.52**
CWP				***0.79***

Note: Italicised, bold figures in the diagonal cells represent the internal consistency

Table 2 above summarises the correlations between these four clusters of behaviours. Diagonal entries in this table represent the internal consistency reliability estimates for each cluster as obtained from the sample of 898.

The relationships between the four clusters described in the two tables above broadly align with the findings reported by Gruys (1999), Gruys and Sackett (2003), Hollinger and Clark (1983), and Robinson and Bennett (1995) for scales related to organisational deviance or deviant behaviours targeted at the organisation, and interpersonal deviance or deviance targeted at colleagues, co-workers or superiors in the workplace.

For example, good attendance shows a commitment and a respect for the expectations of the organisation as is likely to be set out in the terms and conditions of employment. Poor attendance would therefore be an example of organisational deviance. Interpersonal OCBs would reflect a respect and commitment to co-workers such as maintaining a positive emotional attitude to other members of a team. A consistently negative or aggressive interaction with co-workers would therefore be an example of interpersonal CWB.

Table 2 does show a strong relationship between Time Keeping and Meeting Expectations (r=0.70) and these do broadly correspond to measures of organisational OCB versus organisational CWB (see Table 1 above for content of the related items). Similarly, Working with Others and Coping with Pressure exhibit a strong correlation (r = 0.52) and these broadly correspond to measures of interpersonal OCB

versus interpersonal CWB. Overall, all behaviours included in the SHL definition of dependability are correlated (with an average correlation of 0.47) as would be expected from the co-occurrence view of CWBs (Gruys and Sackett, 2003), and the internal consistency reliability of the sum of all dependability scales summarised in Table 1 is 0.84 (N=898).

The four clusters described above represent the behaviours that DSI was designed to predict and these behaviours, in turn, underpin more or less effective customer service and safer versus less safe workplace behaviours. We will explore these links in more detail a little later in this manual, but first we will explore the role of personality in predicting OCBs and CWBs.

The role of personality in OCBs and CWBs

Driven largely by meta-analytic[2] studies, recent years have seen a wealth of research evidence and clarification of the relationships between the dispositions of individuals, usually measured or analysed in the context of the Big 5 model of personality, and OCBs and CWBs. For example, Judge and colleagues (e.g. Judge and Ilies, 2002; Judge et al., 2002) have shown that both OCBs and CWBs are related to the Big 5, most notably in the context of CWBs to Agreeableness, Conscientiousness and Emotional Stability (see Bartram and Brown, 2005, for a summary).

General relationships between personality and OCBs and CWBs

Judge's results are supported by research going back over twenty years. For example, Berry, Ones and Sackett (2007) found that these three personality constructs (Agreeableness, Conscientiousness and Emotional Stability) consistently predicted both interpersonal deviance and organisational deviance with reported effect sizes (sample weighted correlations) ranging between -0.23 and -0.42 for operational validities (corrected for artefacts such as measurement error in the criterion). Sample weighted averages for uncorrected or observed correlations were between -0.19 and -0.34. For overall deviance (interpersonal and organisational combined) the effect sizes (corrected) were between -0.26 and -0.44.

These results are broadly consistent with those reported from an earlier study by Ones (1993a) in which she found effect sizes of between -0.25 and -0.41 for both integrity tests and personality questionnaires used in screening for CWBs. Salgado (2002) found more mixed results from his meta-analysis using the Big 5 taxonomy, but did find that Conscientiousness and Agreeableness were related to deviant behaviours such as disciplinary problems and organisational rule breaking (both of which can be considered examples of organisational deviance). Sackett and Wanek (1996) reported uncorrected validities against organisational deviance of -0.27 for integrity tests and -0.20 for personality tests (the signs of these estimates have been made negative here to be consistent with the results of other studies reported in this section). In a more recent publication, Ones and Viswesvaran (2001a) report mean corrected (operational) validities of -0.39 and -0.47 respectively for Emotional Stability and Conscientiousness, as well as -0.32 for integrity tests (again, signs have been made consistent with other results in this section). Marcus, Lee and Ashton (2007) also found that Agreeableness, Conscientiousness and Emotional Stability were consistent predictors of CWBs across Canadian and German samples.

Personality and service (customer) orientation

In a much earlier study that presages the approaches reflected in the more recent work just cited, Hogan, Hogan and Bush (1984) explored the attitudes and behaviours influencing the quality of the interactions between an organisation and its customers or clients. Specifically, they looked at service orientation which they saw as applying "... to all jobs in which employees must represent their

[2]Meta-analysis is essentially an analysis of the results reported by other studies, an analysis of analyses. Meta-analytic methods are used to identify whether relationships are consistent across studies and to identify factors (characteristics of studies or other variables that can be identified as associated with studies) that influence the size of the relationships found.

organization to the public and where smooth and cordial interactions are required" (see also the definitions of positive versus poor customer service experiences provided later in this manual). Their research on various public sector positions showed that service orientation was most closely associated to what they referred to as Likeability (Agreeableness) and Adjustment (Emotional Stability) such that those seen as exhibiting stronger service orientation were also more likely to be cooperative, rule following, attentive to detail and not variety seekers, as well as self-controlled, dependable and well-adjusted.

Personality and accidents

Looking at the relationships between workplace accidents and personality constructs, Clarke and Robertson's (2008) meta-analysis found that lower Agreeableness, lower Conscientiousness and lower Emotional Stability were all associated with higher accidents for individuals. They also found that higher Openness to Experience was also associated with higher accidents, a result we shall return to later when we consider the construct validity of DSI. It should be noted that Clark and Robertson caveat their results as being suggestive of relationships with the Big 5 given the relatively small number of studies they were able to identify and include in their analysis.

Digman's higher order factor Alpha

Our review of the research literature (results that cover over 200,000 participants in the various individual studies covered by this published research) clearly suggests that three higher-order constructs of the Big 5 have consistent relationships with CWBs, service behaviours as well as with workplace accidents. The wider literature (e.g. Hunter and Schmidt, 1999) shows that at least one of these, Conscientiousness, is a consistent general predictor of work performance. The Hogan et al. (1984) and the Clark and Robertson (2008) studies also lend support to these three constructs relating to different workplace outcomes; that is, to the co-occurrence hypothesis described earlier in this manual.

These three Big 5 constructs have also been proposed by Digman (1997) as one of the higher-order factors of personality that he has labelled Alpha[3], and which he has defined as a socialisation factor related to impulse restraint, conscience, the management of hostility and aggression, as well as neurotic defence. He makes the distinctions between agreeableness versus hostility, conscientiousness versus heedlessness, emotional stability versus neuroticism, and that those higher on Alpha are likely to exhibit higher impulse restraint and conscience.

One of the criticisms of integrity tests and more overt and empirically driven approaches to screening applicants or employees for CWBs has been the lack of a theoretical framework through which we can understand the antecedents of OCBs and CWBs. However, with the emergence of frameworks such as Digman's that help to explain the co-occurrence of personality constructs in predicting OCBs and CWBs, and with the emergence of general taxonomies of OCB and CWB behaviours, a clearer

[3]Digman's model has two higher-order factors the second of which, Beta, relates to learning and growth and is defined by Extroversion and Openness-to-Experience.

understanding of the relationships between dispositions, behaviours and workplace outcomes is now possible. Indeed, Ones and Viswesvaran (2001a) state that Alpha could "... be the most important trait that needs to be systematically measured among job applicants".

As well as asking managers and supervisors to rate employees on the dependable behaviours described in the previous section, we also asked them to rate employees using a series of reference scales for Agreeableness, Conscientiousness and Emotional Stability. This is a slightly different approach to that generally used in Big 5 research in that the more usual study design is to request employees to self report on Big 5 items and to correlate these with either supervisor/manager ratings, or directly with harder criteria such as absenteeism or attendance. We included these Big 5 constructs in the supervisor/manager ratings for two reasons. The first was to explore the extent to which these three constructs, and therefore Alpha, do influence perceptions of employees and specifically customer service orientation and accident proneness. The second, and subject to Alpha being a significant factor influencing supervisor/manager perceptions of performance, to explore the construct validity of the dependability measures we defined as criteria for validating DSI.

Table 3 overleaf summarises a meta-analysis of the relationships across studies between the four dependability behavioural clusters and the three Big 5 elements of Digman's Alpha using the procedures described by Hunter and Schmidt (2004) for conducting a meta-analysis of correlations. The table is presented in two parts: Part A reports the correlations observed across the 13 validation studies in the DSI programme; Part B reports the variance accounted for once statistical artefacts were included in the analysis. The number of studies (k) is 13, the overall sample is again 898 with an average sample size per study of 69 with sample sizes ranging from 40 to 143. For each pair of variables in Table 3A, the uncorrected sample weighted average correlation is reported first followed by the sample weighted average corrected or operational correlation. Corrections within each study were based on two artefacts: corrections for attenuation or unreliability were based on the internal consistency estimates obtained for each respective scale (e.g. Time Keeping and Conscientiousness) within each study; corrections for range restriction were based on the ratio of variances for scales in each individual study to variances obtained for the overall aggregate sample of 898.

Table 3: Meta-analysis of relationships between ratings of employees on dependable behaviours and Alpha constructs

A: Correlations	Time Keeping (TK)	Meeting Expectations (ME)	Working with Others (WWO)	Coping With Pressure (CWP)
Conscientiousness	0.66*	0.82	0.37	0.54
	0.94**	1.00	0.49	0.70
Agreeableness	0.44	0.61	0.55	0.55
	0.61	0.75	0.71	0.70
Emotional Stability	0.19	0.26	0.48	0.45
	0.29	0.33	0.64	0.58

Note: * uncorrected for statistical artefacts. ** corrected for statistical artefacts.

B: Variance Accounted For	Time Keeping (TK)	Meeting Expectations (ME)	Working with Others (WWO)	Coping With Pressure (CWP)
Conscientiousness	100%	100%	44%	87%
Agreeableness	50%	47%	35%	62%
Emotional Stability	55%	41%	25%	25%

The relationships identified in Table 3 can be summarised as follows:

- Strong and consistent relationships between Time Keeping and Meeting Expectations with Conscientiousness
- A strong and consistent relationship between Coping with Pressure and Conscientiousness
- Though less consistent, the results do indicate a strong relationship between Agreeableness and all four dependability clusters
- The results also indicate a strong relationship between Emotional Stability and Working with Others and Coping with Pressure as might be expected for aspects of workplace behaviour linked to interpersonal CWBs.

Overall, the results presented in Table 3 suggest that the dependability behaviours used in the DSI validations to be reported later in this manual do capture manager and supervisor perceptions that are consistent with Digman's Alpha and related psychological constructs.

Towards a systemic model for predicting workplace outcomes: Linking disposition, dependability, customer service and accident proneness

So far we have looked at a number of components in the promotion of OCBs and the management of CWBs in organisations: behaviours that define dependability, and dispositions that link to OCBs and CWBs as well as specific outcomes in the workplace such as customer service and accidents. In this section, we will describe an integrated model that brings all these elements together to predict outcomes in the workplace based on scores obtained from the DSI.

The model is summarised in Figure 1 below. The model proposes that the likelihood of a positive or a negative outcome in the workplace is influenced by critical workplace behaviours as described by the SHL definition of dependability. Positive or negative aspects of these behaviours exhibited by people working in an organisation or work unit are influenced by a general disposition as measured by DSI. Finally, attributes of candidates or employees can be identified that predict the likelihood of positive or negative aspects of the critical workplace behaviours being exhibited by an individual or work group.

Figure 1: Linking disposition to behaviours and workplace outcomes

This model follows from the approach suggested by Viswersveran (2002) in looking for links between workplace behaviours and workplace outcomes (in effect, models of the causal relationship between criterion measures or different clusters of endogenous or Y variables), and with predictors of the workplace behaviours that have been shown to influence the likelihood of desired organisational outcomes (in effect, different exogenous or X variables). This approach to understanding workplace behaviours is also described by Bartram, Robertson and Callinan, (2002).

DSI operates in the model shown in Figure 1 in much the same way as a Criterion-focused Occupational Personality Scale or COPS as defined by Ones and Viswersveran (2001a and 2001b). That is, and in contrast to traditional approaches to personality scale development, DSI was designed specifically to predict and to be interpreted through the four dependability behaviours. In further contrast to many instruments that claim to be a COPS, the DSI provides a single score that is specifically targeted at and directly reflects the likelihood of the OCB versus CWB aspects of the dependability behaviours being exhibited.

As will be described later in this manual, this likelihood has been classified according to five score bands which enable quick and simple interpretations, as well as providing the user with options for where to set the cut-scores reflecting the levels of risk management they wish to operate. In this way, and given that DSI scores are interpreted in terms of the criterion behaviours and not in terms of an elaborated model of the personality of the individual, DSI can also be considered a competency-based assessment.

Using the model to predict customer service outcomes

Consider a good customer service experience that you have had, and then consider the extent to which you felt that the person who served you focused on your needs as a customer and helped you to make, what you see as, a better purchasing decision.

Now think of a poor customer service experience and the extent to which the person who served you did not focus on your needs and did not help you to make a better purchasing decision (or perhaps you consider that they did as you took your custom elsewhere).

These contrasts reflect the type of items reported by Taylor, Pajo, Cheung and Stringfield (2005) for their scale measuring the extent to which someone is customer focused. We adopted this scale as one of the measures of workplace outcomes completed by supervisors and managers in rating the employees who participated in the DSI validation studies.

Using the model to predict safety outcomes

Consider someone that you know to be more accident prone than others that you know. To what extent do they tend to rush to get things finished, cut corners to get things done and tend to forget to inform people? In exploring the relationship between DSI and accident proneness, we focused primarily on statements that relate to cognitive failure as conceived by Broadbent, Cooper, Fitzgerald and Parkes (1982) as underpinning many accidents. Broadbent and colleagues proposed that there are a set of behaviours that indicate a consistent failure to plan ahead, to allow sufficient time to complete a task, to follow the correct procedures and to keep those involved informed and that these behaviours underpin many accidents.

This, then, is the second workplace outcome that we included in the programme using a scale developed by the first author of this manual, Eugene Burke, and for which we would expect a negative relationship with the dependability behaviours (more dependable related to lower accident proneness).

Using the model to predict overall perceived value of employees

Finally, we asked supervisors and managers to evaluate employees in terms of their overall performance and value to the organisation on a four point scale ranging from below average, average and above average through to outstanding.

The evidence presented in the next two sections of this manual

We will use the model described in Figure 1 to structure the results obtained from the DSI criterion validation studies. First, we will describe the relationships between dependability behaviours and customer service, and between dependability and accident proneness, after which we will then report the validities for DSI in predicting the dependability behaviours across a variety of jobs and work settings.

Dependability and manager and supervisor perceptions of employees

Predicting outcomes in customer service roles

Data were obtained from seven studies and a total of 570 employees working in customer facing roles ranging from car hire (US), care services (UK), customer services in communications industry (UK), consumer retail (UK), hotels (Australia) and video rental (UK). The sample weighted average correlation between overall dependability ratings (all four clusters) and customer service orientation using the scale provided by Taylor et al. (2005) was, uncorrected, 0.70 (with correlations within studies ranging from 0.54 to 0.84) and corrected 0.87 (corrected for measurement error and range restriction). Statistical artefacts accounted for all the variance in corrected correlations suggesting that the relationship between dependability and customer service orientation is generalisable across the work settings and job roles covered by the data.

To explore the relationships between dependability and customer service orientation in more detail, and given the meta-analytic results showing the relationship to be generalisable, all data across studies were aggregated and customer service orientation ratings regressed onto the four dependability cluster scores. The results are shown below in Table 4. The Multiple R shown approximates the sample weighted correlation cited above of 0.7 as the regression was directly computed from the raw data and no correction for measurement error was made in this analysis.

Table 4: Relationships between customer service orientation and dependability clusters (N=570)

Multiple R	0.75	Significance	0.001
Dependability Cluster	Zero Order r	Partial r	Beta Weight
Time Keeping (TK)	0.44	0.05	0.04
Meeting Expectations (ME)	0.64	0.32	0.34
Working with Others (WWO)	0.38	0.04	0.03
Coping with Pressure (CWP)	0.68	0.47	0.46

The results summarised in Table 4 indicate that the relationship between dependability and customer service orientation is primarily driven by the Meeting Expectations and Coping with Pressure clusters, with these clusters having zero order correlations of 0.64 and 0.68 respectively with manager and supervisor ratings of employee customer service orientation[4].

Earlier in Table 3, we observed that the dependability behaviours were found to have moderate to high correlations with constructs defining Digman's Alpha. From the data obtained from the samples reported on in this section, customer service orientation was observed to correlate 0.63 with conscientiousness ratings by managers and supervisors of employees, 0.72 with agreeableness ratings and 0.21 with emotional stability ratings.

[4]Given the moderate to high correlations observed between the four clusters as shown in Table 2, collinearity checks were undertaken and these were found to be within acceptable tolerances.

It is therefore reasonable to suggest that managers or supervisors in customer facing functions rate employees as higher on customer service orientation who are seen as more conscientiousness and agreeable, which manifests in the same employees being more likely to meet organisational and role expectations (i.e. not exhibit one facet of organisational deviance) and being able to cope with day-to-day work challenges (i.e. less likely to manifest one facet of interpersonal deviance).

Predicting outcomes in safety critical roles

Data were obtained from six studies and a total of 328 employees working in safety critical roles ranging from construction workers (primarily delivery truck drivers and loaders in the US), engineers (two separate samples from Australia and the aviation industry, and from defence in the UK), manufacturing (US), mining (South Africa) and train drivers (UK). The sample weighted average correlation between overall dependability ratings (all four clusters) and ratings of accident proneness (cognitive failure) was, uncorrected, -0.49 (with correlations within studies ranging from –0.23 to -0.74) and corrected -0.63. Statistical artefacts accounted for 27% of the variance in corrected correlations suggesting that the relationship may be subject to moderators. However, the lower bound 90% credibility estimate was still substantially below zero at -0.38 suggesting that the effect of moderation is in the size of the relationship rather than the direction or substance of the relationship between dependability and accident proneness.

Table 5 summarises the results of regressing accident proneness ratings onto the four dependability clusters. The Multiple R obtained approximates the uncorrected weighted average correlation observed between dependability and accident proneness as reported above. The results of the regression analysis suggest that the strongest relationships were found for Time Keeping and Meeting Expectations. In other words, accident proneness is most strongly related to the two clusters representing aspects of organisational deviance.

Table 5: Relationships between accident proneness (cognitive failure) and dependability clusters (N=328)

Multiple R	0.51	Significance	0.001
Dependability Cluster	Zero Order r	Partial r	Beta Weight
Time Keeping (TK)	-0.45	-0.24	-0.27
Meeting Expectations (ME)	-0.46	-0.24	-0.31
Working with Others (WWO)	-0.15	-0.074	0.04
Coping with Pressure (CWP)	-0.24	0.00	0.00

The three reference scales used as markers for Digman's Alpha were found to correlate respectively with accident proneness as follows: manager/supervisor ratings of conscientiousness -0.42, ratings of agreeableness -0.44 and ratings of emotional stability -0.43 (all correlations are uncorrected for artefacts).

Taken together, these results would suggest that those who are seen by managers and supervisors as more accident prone are also seen as less conscientious, less agreeable and less emotionally stable. These behaviours might be manifested in not checking for errors and not sticking to company regulations (i.e. organisational deviance as represented by the CWB aspect of Meeting Expectations), as well as poor time management and attendance (i.e. organisational deviance as represented by the CWB aspect of Time Keeping).

A summary of the relationships between dependability and workplace outcomes

Figures 2 and 3 below summarise the relationships identified between overall ratings on the four dependability clusters, and customer service and accident proneness respectively. These figures also place the results in the context of the performance model outlined in Figure 1. That is, positive behaviours as defined by the four dependability clusters are highly related to and influence levels of customer service and accident proneness as rated by experienced managers across a range of workplace settings. Note that the first value shown above the directional arrows in each figure is the operational correlation obtained from meta-analysis and after corrections for statistical artefacts, while values in parentheses are the results from the regression analyses reported earlier in this section and are uncorrected for artefacts.

Figure 2: Relationship between customer service orientation and dependability

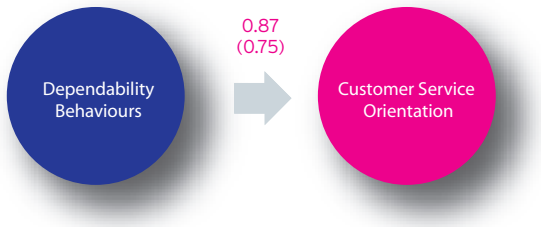
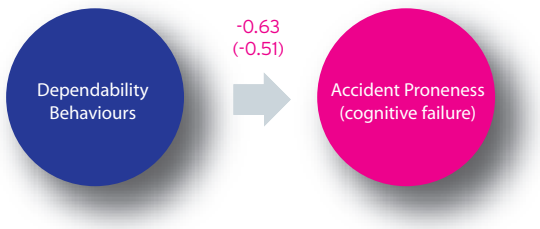


Figure 3: Relationship between accident proneness (cognitive failure) and dependability



As we will see in the next section of this manual, the relationships between DSI and the four dependability clusters, as well as dependability overall, are generalisable irrespective of whether the job setting is customer facing or safety critical. There are some variations in the strength of the relationships found for accident proneness, as we have seen above. As noted there, these variations represent differences in the strength of the relationship rather than its presence or its direction.

For completeness and as a prelude to evidence of DSI's criterion validity, we present the correlations observed across all major criteria for the sample of 898 and the 13 validation studies in Table 6. All criteria that have been discussed in this section were included in all studies irrespective of whether they were customer facing or safety critical. It may interest the reader that the correlation between customer service orientation and accident proneness was observed to be -0.43 (uncorrected for measurement error) and -0.51 when corrected for unreliability (attenuation) in the respective scales.

Table 6: Observed correlations between criterion constructs as rated by managers for 898 employees across 13 studies (computed on aggregate data)

Criterion Measure	Customer Service Orientation	Accident Proneness	Overall Rating of Performance
Time Keeping (TK)	0.69	-0.49	0.58
Meeting Expectations (ME)	0.68	-0.50	0.64
Working with Others (WWO)	0.72	-0.46	0.45
Coping with Pressure (CWP)	0.24	-0.43	-0.24

The construction of DSI and evidence supporting its criterion validity

In this section, we will focus on the relationship between DSI and the dependability behaviours as shown on the left handside of Figure 1 (p.19) and the model of performance used to develop DSI.

The construction and scoring of DSI

Earlier, DSI was positioned in the performance model as analogous to a Criterion Oriented Personality Scale or COPS following the definition offered by Ones and Viswersveran (2001a and 2001b). The focus in developing DSI was to provide a short, fake resistant instrument that could be used as an efficient screening tool or as one assessment component in combination with other assessments for use in selection of operational personnel.

A key design aim in developing DSI was to construct items, made up of statement pairs, using the following logic and exhibiting the following features:

- Each pair to contain one statement keyed as either a positive or a negative predictor of dependability
- Each pair to contain one statement operating as a distractor (i.e. not hypothesised as a predictor of dependability)
- Both statements in each pair to be matched in terms of attractiveness, (i.e. seen as a desirable characteristic of people by those completing DSI)
- A simple response format where the respondent indicates which statement (out of option A and B) is 'most like' them, or indicates that neither statement applies to them, or that both statements are equally applicable.

Over the past five years, SHL has completed a number of mappings and correlational studies between the Occupational Personality Questionnaire (OPQ) and the Big 5 personality constructs (Bartram and Brown, 2005). Using these mappings as a guide, content in the Work Styles Questionnaire (SHL, 1999) was sampled as the basis for constructing statement pairs to tap into facets of personality related to dependability.

This design aim had as its goal a quasi ipsative or forced choice structure which would be appropriate for lower levels of education through ease and simplicity of response format. This would also operate to reduce faking or false responding (see Jackson and Wroblewski, 2000, and Christiansen, Burns and Montgomery, 2005, for general reviews of the use of ipsative formats to reduce faking on self-report questionnaires). Since each statement pair has three possible response options and given that Version 1.0 of DSI contains 22 statement pairs, there are 3^{22} possible response permutations to the DSI which reduces the ability of applicants to guess the correct answers. Version 1.1 of DSI has 18 statement pairs, therefore there are 3^{18} possible response permutations.

Data from 303 paper-and-pencil assessments using the WSQ in the UK and US were analysed to identify the attractiveness of WSQ items. Attractiveness was defined as a high endorsement of an item. Items were then classified into positive, negative or distractor item categories matched in terms of attractiveness, and then used to construct the statement pairs. Each statement pair was scored on the basis of their hypothesised relationship with dependability at work. Each pair provided a score of 1 (lower dependability predicted), 2 (moderate dependability predicted) or 3 (higher dependability predicted). The overall score was defined as the sum of the scores across all statement pairs. This scoring method was used to evaluate each statement pair to determine whether the pair would be selected for the final instrument, required revision or would be rejected.

Two rounds of pilot studies were conducted. The first pilot study involved operational staff in SHL (N=36 and included administrative, catering and maintenance staff) to test readability and ease of comprehension of the statements. The second pilot study involved a larger sample of 105 employees from client organisations in distribution, waste management and public transportation to test more thoroughly whether the statements functioned as hypothesised. Five statement pairs were rejected in the course of these pilots and 10 pairs were revised to produce the 22 statement pairs used originally and deployed as Version 1.0 of DSI.

It is important to note that the scoring of DSI involves three steps:

- The scoring of each statement pair
- The summing of the statement pair scores
- The transformation of the summed score into an indicator of dependability at work

This three-step process was developed to simplify the interpretation of the DSI score by making a direct reference to behaviours in the workplace, namely the four dependability clusters described earlier in this manual. The transformation into risk bands is described later in this manual.

Revision of DSI and Version 1.1

Detailed evaluations of DSI items that led to the reduction in length of the instrument are described in the later section of this manual on the reliability and fairness of DSI. In summary and to provide a context for the evidence to follow next on criterion validity, item level analyses by gender, age, ethnicity and English language fluency identified four DSI items (each made up of statement pairs) as performing inconsistently and therefore were not adding substantially to the criterion validity of the overall score. These items were therefore removed and DSI Version 1.1 now offers a more efficient screening and assessment tool with no loss of discrimination between candidates or loss of validity.

The 22 and 18 statement pair versions of DSI were observed to correlate **0.95** (N=6,095, as obtained from assessments using Version 1.0), indicating a high degree of consistency in the rankings offered by both versions. A comparison of the sample weighted validities for predicting overall dependability ratings for criterion data available at the time of reviewing the length of DSI showed no loss in

criterion validity due to the removal of four statement pairs. For 5 studies and an overall sample size of 320 (almost equally split across customer facing and safety critical jobs), the 22 statement pair version yielded a sample weighted criterion validity of 0.24 in contrast to 0.28 for the 18 statement pair version (both correlations uncorrected for artefacts of range restriction).

A meta-analysis of DSI criterion validity

Reference has already been made to the 13 studies and 898 participants in validation studies across a number of settings completed by the end of 2008. We will now describe those studies in more detail and provide the results of a meta-analysis evaluating the consistency of DSI predictions. The principal data collection design was a concurrent validation involving existing employees completing DSI and, where participating companies permitted, demographic information providing details of gender, age, work experience and education as well as length of employment with the organisation and service in their current role. Where the SHL demographics questionnaire was not administered, we requested demographics from the participating organisation.

The supervisors or managers of these employees were asked to rate them on the dependability behavioural items, the three Alpha constructs (Conscientiousness, Agreeableness and Emotional Stability), and on the customer service orientation and accident proneness items. Supervisors and managers were also asked how long they had known and been responsible for the employees they were rating. Those employees with less than six months service in their current role and who had been supervised for less than six months by the supervisor/manager completing the criterion ratings were excluded from the analysis on the basis that ratings would reflect lower familiarity with either the job (from the employee's perspective) or the employee's performance (from the supervisor/manager perspective).

Table 7 provides a summary of the characteristics of the seven studies completed for customer facing roles with 570 employees, and six studies completed for safety critical roles with 328 employees. The table has been split into two parts, A and B, to reflect the different settings in which the studies were conducted.

The demographics for each study have been coded to reflect widely used equal opportunities classifications such as male and female, under and over 40 years of age, and white and non-white. We have added an educational split between those with no formal educational qualifications up to a certificate in secondary education (for example, the General Certificate of Secondary Education or GCSE that is generally awarded at age 16 in the UK), and those with formal educational qualifications at baccalaureate (or high school degree in the US and Advanced Level Certificate in the UK as would be generally awarded at age 18).

As Table 7 shows, data were collected primarily from the UK but also in Australia (2 studies), South Africa (1 study) and in the US (3 studies). Generally, females tended to occupy the customer facing roles, and males occupied safety critical roles predominantly. With the one exception of the South African mining study, most roles were occupied by whites as classified by national ethnicity codes. Age distributions did vary by study, but the majority age group across studies was 39 years or younger. We will revisit the relationship between demographics and DSI scores in the section of this manual that explores the reliability and fairness of DSI scores.

Table 7: Characteristics of the DSI criterion validation studies completed between 2004 and 2008

A: Customer facing roles: Total N=570

Country	Job/Role	N	Gender	Age	Ethnicity	Education
Australia	Various in hotel (e.g. concierge, front desk)	52	56% male	78% 39 years or younger	92% white	71% baccalaureate (high school degree) or higher
UK	Public sector care workers	63	97% female	Not provided	Not provided	77% secondary certificate of education or no formal qualifications of education
UK*	Customer service in banking	143	52% male	89% 39 years or younger	89% white	53% baccalaureate (high school degree) or higher
UK	Customer service in telecommunications	63	63% female	76% 39 years or younger	Not provided	66% baccalaureate (high school degree) or higher
UK	Shop retail	78	78% female	51% 40 years or older	99% white	71% secondary certificate of education or no formal qualifications
UK	Video outlets assistants	56	71% male	98% 39 years or younger	92% white	70% baccalaureate (high school degree) or higher
US*	Car hire outlet assistants	115	83% female	59% 40 years or older	66% white	Not provided

Note: * indicates a predictive validation study in which DSI was administered at the recruitment stage and the performance of successful candidates was then followed up post hire.

B: Safety critical roles: Total N=328

Country	Job/Role	N	Gender	Age	Ethnicity	Education
Australia	Aviation engineering apprentices	72	97% male	100% 39 years or younger	Not provided	Not provided
South Africa	Mining operatives	40	89% male	92% 39 years or younger	83% non-white	98% baccalaureate (high school degree) or higher
UK	Navy engineering apprentices	52	100% male	100% 39 years or younger	92% white	71% secondary certificate of education or no formal qualifications
UK	Train drivers	21	90% male	79% 39 years or younger	86% white	69% secondary certificate of education or no formal qualifications
US	Manufacturing machine operators	64	97% male	64% 39 years or younger	70% white	Not provided
US	Construction drivers and loaders	79	100% male	72% 40 years or older	81% white	95% secondary certificate of education or no formal qualifications

Table 8 summarises the results of applying the Hunter-Schmidt model of meta-analysis (Hunter and Schmidt, 1990) to the validity coefficients obtained from these studies. The results are split as per Table 7. The analyses generally showed that predictions of the dependability behaviours were consistent irrespective of the category into which a study fell. This is shown in the third part of Table 8, 8C, which reports results across all 13 studies.

Table 8: Results from meta-analysis of DSI validities by criterion

A: Customer facing roles: Total N=570						
Criterion	**r**	**Range of r**	**SD_T**	**P**	**SD_T**	**Lower credibility value**
Overall Dependability	0.27	0.17 to 0.41	0	0.47	0	Not applicable
Time Keeping	0.26	0.18 to 0.33	0	0.38	0	Not applicable
Meeting Expectations	0.28	0.18 to 0.43	0	0.37	0	Not applicable
Working with People	0.22	0.14 to 0.32	0	0.29	0	Not applicable
Coping with Pressure	0.12	-0.01 to 0.30	0	0.16	0	Not applicable
Conscientiousness	0.29	0.25 to 0.42	0	0.40	0	Not applicable
Agreeableness	0.21	-0.03 to 0.48	7%	0.27	52%	0.23
Emotional Stability	0.11	-0.05 to 0.20	0	0.14	0	Not applicable

B: Safety critical roles: Total N=328						
Criterion	**r**	**Range of r**	**SD_T**	**P**	**SD_T**	**Lower credibility value**
Overall Dependability	0.22	0.16 to 0.29	0	0.38	0	Not applicable
Time Keeping	0.20	0.04 to 0.44	0	0.30	0	Not applicable
Meeting Expectations	0.18	0.003 to 0.46	0	0.23	0	Not applicable
Working with People	0.16	0.009 to 0.3	0	0.23	0	Not applicable
Coping with Pressure	0.13	-0.02 to 0.21	0	0.18	0	Not applicable
Conscientiousness	0.19	0.05 to 0.50	0	0.23	53%	0.17
Agreeableness	0.10	-0.16 to 0.29	0	0.14	19%	Not applicable
Emotional Stability	0.15	-0.15 to 0.24	0	0.21	0	Not applicable

C: All roles: Total N=898						
Criterion	**r**	**Range of r**	**SD_T**	**P**	**SD_T**	**Lower credibility value**
Overall Dependability	0.26	0.16 to 0.41	0	0.44	0	Not applicable
Time Keeping	0.24	0.04 to 0.44	5%	0.36	0	Not applicable
Meeting Expectations	0.24	0.003 to 0.46	30%	0.32	21%	Not applicable
Working with People	0.20	0.009 to 0.32	0	0.27	0	Not applicable
Coping with Pressure	0.12	-0.02 to 0.32	0	0.17	0	Not applicable
Conscientiousness	0.24	0.05 to 0.50	22%	0.33	36%	0.30
Agreeableness	0.16	-0.16 to 0.48	50%	0.21	44%	0.18
Emotional Stability	0.13	-0.15 to 0.24	0	0.17	0	Not applicable

Notes: r = sample weighted uncorrected validity. Range of r = range of observed validities.
SD_T = variance after sampling error. p = operational validity

In most cases, the results shown in 8C show that the validities are generalisable across studies, settings and job roles, as well as geographies. In the cases where true variance (SD_T or the variance remaining after either sampling error and/or statistical artefacts have been accounted for) is substantial across operational validities, the results show that splitting the data into customer facing and safety critical roles does not reduce SD_T substantially or consistently (e.g. the SD_T of 44% in Part C for DSI predictions of Agreeableness ratings by supervisors or managers does drop to 19% for safety critical roles, but increases to 52% for customer facing roles). This may indicate that there are moderators operating that influence the results across studies for one or two of the criterion measures. However, the nature of this influence is likely to be the strength rather than the presence or direction of the relationship between DSI and criteria. Indeed, in no case does the lower credibility limit include zero, suggesting that, while the nature of the relationship may vary for some criteria, the general relationships hold across studies and settings.

The results shown in Table 8 also compare favourably with the results that have been reported in the general literature for Big 5 measures (e.g. Judge, 2002a and 2002b; Ones and Viswersveran, 2001a and 2001b), particularly the operational validity of 0.44 obtained for predictions of the sum of ratings across dependability behavioural clusters. It should be noted that the results in the wider literature tend to be for full personality scales and in many cases scores that represent composites across several scales to provide a Big 5 score, while DSI Version 1.1 comprises 18 statement pairs that take only a few minutes to complete.

Looking at specific criterion measures, the results suggest a stronger relative relationship between DSI and ratings of Time Keeping and Meeting Expectations than with Coping with Pressure. This is mirrored in the stronger relationships with managers' ratings of employees on Conscientiousness (0.32) and Agreeableness (0.21), than with Emotional Stability (0.17). As such, these results suggest that predictions of workplace outcomes offered by DSI will tend to operate primarily through manifestations

of conscientious and agreeable behaviours. This will be reflected in higher versus lower levels of compliance with organisational rules and expectations, or alternatively, higher DSI scores are likely to be reflected in lower organisational deviance.

We will explore construct validity data that helps to explain how and why DSI works in predicting OCBs versus CWBs in a later section. We will now conclude this section with two client case studies that show the value offered by DSI in screening for organisational deviance.

The case of unauthorised absence and customer care service advisers in the energy industry

This and the next case study were obtained from client evaluations of DSI using hard criteria such as absenteeism and accidents. This first case study was undertaken in 2007 and involved 136 customer service advisers for a UK client in the energy industry (gas and electricity supply). Their DSI scores were compared to absences during 2007 as shown in Table 9 below. In this case study, the odds were 1 in 2 as to whether an employee would record an absence during the period covered by the study. Analysis showed that those falling into the lowest 30% of DSI scores were 2.5 times more likely to have 1 or more absences compared to the average employee, and to be 5 times more likely to have I or more absences than those scoring in the top 70% of scores on DSI (bandings were based on general distributions of DSI scores and not on the distributions for this particular client).

Table 9: Comparison of the odds of recording an absence broken down by DSI score

DSI Score Band	Zero absences (A)	1 or more absences (B)	Odds (A : B)
Lowest 30%	18%	82%	1 : 5
Highest 70%	41%	58%	1 : 1
All employees	39%	61%	1 : 2

The case of security guards, absenteeism, accidents and incidents of attacks

The second case study was for Group 4 Security (G4S) in the UK and involved 72 drivers (Burke, Fix and Grosvenor, 2008). Records for drivers were available over six months covering unauthorised absences, vehicle accidents for which they had been responsible and attacks that they and their teams had been victims of. Tables 10A through to 10C summarise the results obtained and these show that:

- Guards scoring in the lowest 30% of DSI scores were **more than twice as likely to record an unauthorised absence** than the average employee
- Guards scoring in the lowest 30% of DSI scores were almost **four times more likely to be responsible for an accident with a company vehicle** than the average employee
- Guards scoring in the top 30% of DSI scores were **2 times less likely to be involved in an attack** than the average employee

Table 10: Absenteeism, accident and attack rates for security guards broken down by DSI scores

A: Unauthorised Absenteeism	Zero absences (A)	1 or more absences (B)	Odds (A : B)
Lowest 30%	80%	20%	4 : 1
Highest 70%	91%	9%	10 : 1
All employees	90%	10%	10 : 1

B: Vehicle Accidents	Zero accidents (A)	1 or more accidents (B)	Odds (A : B)
Lowest 30%	80%	20%	4 : 1
Highest 70%	96%	4%	24 : 1
All employees	95%	5%	19 : 1

C: Attacks	Zero attacks (A)	1 or more attacks (B)	Odds (A : B)
Lowest 70%	59%	41%	1 : 1
Highest 30%	85%	15%	6 : 1
All employees	74%	26%	3 : 1

Further data available from this study suggests that, as indicated by the Future Foundation (2004) survey of errors in the workplace, these statistics may represent a significant blind spot amongst supervisors. Table 11 shows the correlations between supervisors appraisals of these drivers and DSI scores. The reader will note some clear gaps in the relationships between supervisor appraisals and DSI for absences and attacks (all correlations are uncorrected for measurement errors or other artefacts).

Table 11: Correlations suggesting a blind spot among security guard supervisors perceptions of drivers

	Unauthorised Absenteeism	Vehicle Accidents	Attacks
Supervisor Appraisal	0.04	-0.24*	0.03
DSI Score	-0.20*	-0.23*	-0.20*

Understanding why DSI works: Evidence of construct validity for DSI scores

We have already explored the relationships between the dependability behaviours and Big 5 constructs as rated by managers or supervisors. The purpose of this section is to place the DSI scores in the broader context of relationships with other predictor scores and measures of personality. More specifically, to report results from correlational studies using the Work Styles Questionnaire (WSQ; SHL, 1999), the Occupational Personality Questionnaire 32 (OPQ32; Bartram, Brown, Fleck, Inceoglu and Ward, 2006) and Customer Contact Styles Questionnaire (CCSQ; Baron, Hull, Janman and Schmidt 1997).

The availability of data from three separate studies across three extensively validated questionnaires avoids the potential overlap between the WSQ and DSI, as DSI content was originally drawn from WSQ items (though these were revised and new content added in the course of DSI's development). The OPQ32 data also allows DSI scores to be evaluated against Big 5 constructs. Equations validated against personality questionnaires based on the Big 5 structure are available for OPQ32, which enable OPQ32 scale scores to be converted to Big 5 indicators as described by Bartram and Brown (2005). The CCSQ study provides data for evaluating DSI against CCSQ scales.

Automotive engineers and the relationship between DSI scores and WSQ scales

Data were obtained from 65 apprentice engineers employed by a local dealership in South Africa of a major international luxury car manufacturer. The sample was 99% male, all 39 years of age or younger, and the modal educational level of this sample (83%) was advanced vocational qualifications (as might be expected given the context of the study).

Data were available for DSI Version 1.1 scores and for the WSQ, a personality questionnaire designed for use with operational jobs. Based on the review of the research literature described earlier in this manual, relationships were explored between DSI scores and six WSQ scales. These scales are described in Table 12 from which DSI scores can be seen to be positively correlated with WSQ scales Considerate, Dependable, Forward Thinking and Resilient, and negatively correlated with the WSQ scales Decisive and Innovative. Overall, the Multiple R obtained from regressing DSI onto these six scales was 0.57, significant at the 0.001 level. Adjusting for the average reliability of all scales in the regression including DSI, the corrected (construct level) correlation between DSI scores and the composite of the six WSQ scales is estimated to be 0.72.

From the scale descriptors provided in Table 12 and consistent with the Hogan et al. (1984) and Clarke and Robertson (2008) papers, these results suggest that higher DSI scorers are less impulsive and more considered in their responses to situations and to others; while lower DSI scorers are more likely to respond impulsively to events. These results are also consistent with Digman's definition of Alpha.

Table 12: Relationships between DSI scores and WSQ scales (N=65 automotive engineer apprentices, South Africa)

WSQ Scale	Higher score definition	Lower score definition	Beta weight from regression
Considerate	Shows consideration; patient; sympathetic; sensitive to others	Tends to be a little insensitive and unsympathetic to others	0.28
Dependable	Hardworking; conscientious and trustworthy; perseveres with routine tasks	May be less conscientious than colleagues; more likely to cut corners and bend rules	0.26
Forward Thinking	Prepares well in advance; plans and organises work; likes structure	Tends to deal with problems as they arise; spends little time planning or preparing in advance	0.23
Resilient	Calm; steady under pressure	Tends to be less relaxed; more anxious; more apprehensive about future events	0.24
Decisive	Likes to resolve problems quickly; jumps to conclusions; impatient; maybe impulsive	Prefers to think things through carefully; reserves judgment until options have been considered	-0.38
Innovative	Comes up with ideas and novel solutions; creative; looks for new ways of doing things	Tends to adopt straightforward and predictable solutions to problems	-0.22

OPQ32 and the relationship between DSI scores and Big 5 indicators

Data were obtained from 427 applicants to a major public sector employer in South Africa. The sample were all 39 years of age or younger (51% were between the ages of 21 and 24 years), 63% were male and the majority were educated to graduate or postgraduate level (80%), and 89% of the sample were Black African.

Data were available for DSI Version 1.1 and OPQ32 with all instruments administered in English. OPQ32 scores were transformed into Big 5 indicators using equations developed by Bartram and Brown (2005) based on structural equation modelling of OPQ32 and Big 5 reference questionnaires. DSI scores were regressed on the OPQ32 Big 5 scores yielding a Multiple R of 0.41 significant at the 0.0001 level. Adjusting for the average reliability of all scales in the regression including DSI, the correlation between DSI scores and a composite of Big 5 scores as weighted by the results of the regression model, the corrected correlation is estimated to be 0.54.

The results of this regression analysis are shown overleaf in Table 13. These show DSI scores to be positively (and significantly) related to Conscientiousness, Agreeableness and Emotional Stability, but negatively (and significantly) related to Openness-to-Experience. These results bear a strong resemblance to the validities reported by Clarke and Robertson (2009) for Big 5 constructs in predicting accidents.

Table 13: Relationships between DSI scores and Big 5 (OPQ32) scores (N=427 applicants to public sector organisation, South Africa)

OPQ32 Big 5 indicator	Zero order correlation	Beta weight from regression
Conscientiousness	0.31	0.27
Agreeableness	0.22	0.18
Emotional Stability	0.17	0.13
Openness-to-Experience	-0.14	-0.19
Extroversion	0.08	0.00

Note: Big 5 indicators (scores) obtained from modelling reported by Bartram and Brown (2005)

The data made available from this study also allows the relationship between DSI scores and SHL's Universal Competency Framework (UCF) to be explored. Table 14 shows the correlations between UCF competency potential scores as obtained from OPQ32 (see Bartram et al., 2006, and Burke, 2008, for further information on the UCF and the function of OPQ32 in providing measures of potential against this framework).

The results shown in Table 14 support DSI as a measure of potential for roles involving observance of organisational values and policies (see the correlations with potential against UCF dimensions 2.2 and 6.1); where planning and quality are important (see the correlations with potential against UCF dimensions 6.2 and 6.3); and where creativity and adapting to change are less important (see the correlations with potential against UCF dimensions 5.2 and 7.1). The results also show that DSI can sit alongside other measures of fit for a role or job given the near zero correlations observed with the remaining 14 UCF dimensions. As such, DSI offers an efficient pre-screen prior to more detailed assessments for fit for a role, or can operate as an efficient component in a broader set of assessments of job/role fit.

Table 14: Relationships between DSI scores and UCF dimensions (N=427).

UCF Dimension	Zero order correlation with DSI
UCF 1.1: Deciding & initiating action	-0.09
UCF 1.2: Leading & supervising	0.07
UCF 2.1: Working with people	0.09
UCF 2.2: Adhering to principles & values	**0.24**
UCF 3.1: Relating & networking	-0.09
UCF 3.2: Persuading & influencing	-0.04
UCF 3.3: Presenting & communicating information	0.06
UCF 4.1: Applying expertise & technology	0.07
UCF 4.2: Analysing	0.02
UCF 4.3: Writing & reporting	0.05
UCF 5.1: Learning & researching	0.06
UCF 5.2: Creating & innovating	**-0.19**
UCF 5.3: Formulating strategies & concepts	0.01
UCF 6.1: Following instructions & procedures	**0.29**
UCF 6.2: Delivering results & meeting customer expectations	**0.31**
UCF 6.3: Planning & organising	**0.32**
UCF 7.1: Adapting & responding to change	**-0.25**
UCF 7.2: Coping with pressure & setbacks	0.02
UCF 8.1: Achieving personal work Goals & objectives	0.09
UCF 8.2: Entrepreneurial & commercial thinking	-0.01

Note: Lines in bold show correlations significant at the 0.01 level.

International bank call centre and the relationship between DSI and the Customer Contact Styles Questionnaire (CCSQ)

Data were obtained from 429 applicants for call centre positions (inbound and outbound) for a large international bank working within the UK. The demographics for this sample were 62.9% female; 88.3% between 16 years and 30 years of age, with the age range extending to 60 years of age; 66.7% identified themselves as White European, 21.4% as Eurasian, 6.1% as Black, 1.9% as Asian and 4% as Other.

In addition to DSI scores, data were available from other assessments including cognitive ability tests (Verify Verbal and Numerical Reasoning, which we will explore a little later in this manual) and CCSQ. The relationships for DSI and CCSQ scales shown in Table 15 are consistent with relationships identified between DSI and the other personality instruments, WSQ and OPQ. The CCSQ scales shown are those identified from regression modelling to be those contributing substantially and significantly to the prediction of DSI scores from the CCSQ scales (Multiple R of 0.49 for all scales, fully saturated model, and 0.48 for the model with just the scales shown in Table 15, restricted model). Corrected for unreliability in instruments, the relationship between DSI and a composite of the scales shown in Table 15 is estimated to be 0.62 at the construct level (i.e. adjusted for measurement error in DSI and CCSQ scales).

Those who score higher on scales of Self Control and Resilience on CCSQ (related to the impulse control aspect of Digman's Alpha); who score higher on Detail Conscientious and Conscientiousness but lower on Flexibility and Innovative (related to the conscientious versus heedlessness aspect of Digman's Alpha, as well as the relationships described earlier between Openness to Experience from the Big 5 and accidents), and who score higher on Participative but lower on Competitive (related to the Agreeableness aspect of Digman's Alpha) score higher on DSI.

Table 15: Regression of DSI scores on CCSQ Scales (N=429).

CCSQ Scale	Standardised (Beta) Weight (observed)	Standardised (Beta) Weight (corrected for measurement error)
CR2: Self Control	0.20	0.26
CR5: Participative	0.22	0.29
CT2: Innovative	-0.12	-0.15
CT3: Flexible	-0.10	-0.13
CT5: Detail Conscious	0.16	0.21
CT6: Conscientious	0.10	0.13
CE1: Resilience	0.13	0.17
CE2: Competitive	-0.21	-0.27

Relationship between DSI and cognitive ability test scores

The study just described also provided data on the relationship between DSI scores and scores on cognitive ability tests, namely verbal and numerical online tests from the Verify Range of tests (Burke, van Someren and Tatham, 2006). Scores on the verbal and numerical test were combined with equal weight to provide an overall estimate of general mental ability, and this composite score yielded a correlation of -0.04 with scores on DSI. Essentially, this data suggests that DSI is uncorrelated with cognitive ability which, as will be discussed later in this manual in relation to research on faking on self-report questionnaires, also shows that score profiles for those with higher general mental ability levels are similar to those at lower levels of the general mental ability range.

Setting DSI score bands to provide levels of risk management in screening potential employees

In the original development work reported in the technical manual for Version 1.0 of DSI (Burke and Kirby, 2006), a series of analyses showed that DSI scores could be used to predict levels of risk of appointing someone into a customer facing or a safety critical role. These analyses, using logistic regression (see Dwyer, 1983, for an introduction), essentially provided an algorithm that had an exponential relationship between DSI scores and effective customer service orientation or lower proneness to accidents. In developing Version 1.1, we have sought a simpler method for classifying DSI scores that retains the exponential relationship with workplace outcomes. This method also provides sufficient scope for other assessments used to identify specific job or role fit to be deployed alongside or subsequent to an administration of DSI.

The breakpoints used for the risk bands described in more detail below were obtained from a sample of 6,095 live administrations of DSI with scores on DSI Version 1.0 equated to DSI Version 1.1 using equipercentile equating (Kolen and Brennan, 2009). Table 16 provides a summary of available demographics for this sample. The distribution of scores used had a mean of 42.84 and an SD of 6.52. Jobs levels in this sample included unskilled or semi-skilled jobs such as production workers, construction workers, baggage handlers, drivers and customer service roles in retail and finance, call centre roles, and extended to skilled technical jobs such as apprentice engineers in heavy engineering, aviation and the automotive industry.

Table 16: Summary of demographics (N = 6,095)

Demographic/Firmographic	
Country	4% Australia, 2% South Africa, 70% UK and 24% US
Gender	55% male
Age	Range 18 to 64 with 64% between 21 and 34
Education	Range from no formal educational qualifications to postgraduate studies with 69% attaining qualifications between certificate of secondary education to high school diploma

The risk bands associated with DSI Version 1.1 using the distribution of scores just described are as follows:

- Very High risk as represented by scores that fall into the lowest 10% the DSI score distribution
- High risk as represented by scores falling into the next 10% of the distribution of DSI scores
- Moderate risk as represented by scores falling into the next 15% of the distribution of DSI scores
- Moderate to low risk as represented by scores falling into the next 15% of the distribution
- Low risk as represented by the top 50% of the distribution of DSI scores.

The work reported in the manual for Version 1.0 clearly showed that there was a threshold at about the median DSI score above which the risk of poor customer service orientation or higher accident proneness did reduce significantly. By allowing for a broader low risk margin, and as mentioned above, there is sufficient scope for other assessments such as questionnaires, tests and interviews to evaluate the fit of the individual to more specific job/role requirements. As such, DSI offers the user the facility to screen for risk and to select for fit, thereby managing the costs of recruitment, minimising the impact of CWBs among new hires as well as maximising the return on the investment in recruitment and selection by ensuring person-job and person-organisation fit.

Figure 4 provides a summary of the risk bands associated with DSI as classified by a red-amber-green (RAG) coding. The descriptions offered for each band of risk emphasise the function of DSI in terms of fit to specific types of roles and environments. For example, where shift patterns and time attendance are important to effective operations in the workplace; where observance of company policies and procedures are important such as in the case of safety critical roles; and where team working is also an important factor.

Figure 4: Summary of DSI risk bands

Band	Interpretation	**Likely Impacts** (for work in general)
✓✓	Low Risk	A **low risk** candidate is likely to have a **strong fit** to jobs where step-by-step procedures, team working and strict working hours are important
✓	Moderate to Low Risk	A **moderate to low risk** candidate is likely to have a **reasonable fit** to jobs where step-by-step procedures, team working and strict working hours are important
●	Moderate Risk	A **moderate risk** candidate is likely to have a **moderate fit** to jobs where step-by-step procedures, team working and strict working hours are important
✕	High Risk	A **high risk** candidate is likely to have a **weak fit** to jobs where step-by-step procedures, team working and strict working hours are important
✕✕	Very High Risk	A **very high risk** candidate is likely to have a **very weak fit** to jobs where step-by-step procedures, team working and strict working hours are important

Reliability and fairness of DSI scores

This section describes the results of studies conducted to establish the stability of DSI scores over time and analyses undertaken to investigate the performance of DSI across different demographic groups. We have linked these topics together in this section as both issues relate to two critical aspects of organisational justice which is seen as important in establishing the credibility of any measure used to support the recruitment and selection of personnel (see Gilliland and Hale, 2005, for more information on dimensions of organisational justice):

- **Procedural justice** relates to whether a process is seen as offering a fair opportunity for participants in that process to demonstrate their suitability for a position or role. The accuracy and stability of an instrument such as DSI allied with strong criterion and construct validity evidence. As already described for DSI in this manual these are critical elements of scientific evidence in supporting positive perceptions of procedural justice. Evidence that shows that an instrument functions equally well for different demographic groups and that it is free from any biases in its content and scoring is also important in supporting positive perceptions of procedural justice.
- **Distributive justice** relates to whether the outcomes of a process such as decisions to hire or not to hire someone are seen as fair. We have conducted extensive analyses across different demographic groups to identify how DSI is likely to perform when different cut-points are applied. For example, we have applied the 4/5th's rule as used in the US to evaluate whether a process or stage of a process may exhibit adverse impact against protected groups as defined by US employment laws (similar classifications are used in other nationalities, but we have used the 4/5th's rule as it is widely used in countries other than the US).

The reliability of DSI scores

Reliability estimates provide information on the consistency and accuracy of scores obtained from a test. Reliability can be estimated in different ways depending on the question being asked:

- To answer the question of how a test score is affected by the quality of the items in a test, reliability can be estimated using the Internal Consistency Coefficient. This reports the proportion of variation in scores that can be attributed to consistency (or lack thereof) in the measurement properties of the items in the test. A key assumption for this form of reliability estimate is that the scale from which the score is obtained is unidimensional (i.e. measures a single construct). DSI does not meet this assumption.
- To answer the question of how a test score is affected by variation in the measurement qualities of different versions of a test (i.e. which version is administered to an applicant), reliability can be estimated using the Alternate Forms Coefficient. This reports the percentage of variation in scores that can be attributed to differences across alternate test forms. At present, the DSI does not have an alternate form.
- To answer the question of how consistent scores are over time, then reliability can be estimated by the Test-retest or Stability Coefficient which reports the proportion of variation in applicants' rankings on test scores across two or more administrations at different times.

The DSI was developed in much the same way as a criterion-referenced measure where the focus is on predicting a later outcome. In the case of DSI this is the four dependable behaviours, rather than a unidimensional scale in the more classical model of self-report questionnaire scales. The DSI score is a composite of responses to pairs of statements that have been individually keyed as indicators of dependable behaviours in the workplace. As such, and given that only a single form exists, the most appropriate method of estimating the reliability of the DSI is the stability coefficient.

A sample of 71 people across two offices of a business services company based in the UK and Australia participated in the test-retest trial of DSI with a time gap of 5 to 9 working days between administrations. The sample comprised junior administrative positions up to professional managers. Sixty-three percent of the sample was female with age ranging from 25 to 45. There was a mix of educational backgrounds covering little formal education to graduate and postgraduate degrees. The correlation between first and second DSI scores (the estimated test-retest reliability or stability coefficient) obtained from this sample was **0.72**. This is the reliability estimate used in the meta-analyses reported earlier in this manual where corrections for measurement error associated with the DSI scores were undertaken.

From the reliability estimated for a scale, the standard error of measurement or SEM can be calculated using the formula $(1-r_{xx})^{1/2}$ X SD, where r_{xx} represents the estimated reliability of the scale and SD is the scale standard deviation. The SEM is used to define a range within which a person's true score is likely to lie. The SEM for the DSI is given by $(1-0.72)^{1/2}$ X 6.52, where 6.52 represents the standard deviation of DSI scores for DSI Version 1.1 as obtained from a sample of 6,095 job applicants across Australia, North America, South Africa and the UK. The SEM for DSI in raw score terms is therefore 3.45 or 3 raw score points. For example, if a person obtains a score on the DSI of 50, then there is a 68% chance that the person's true score lies between 47 (1 SEM below the observed score) and 53 (1 SEM above the observed score).

For those who are familiar with Version 1.0 of DSI and who may be users, the correlation between both versions of the instrument is **0.95** (as noted on page 27 of this manual, this correlation is based on a sample of 6,095). As such, there is a high consistency in the scores obtained between the two forms of the instrument. This reflects the removal of four items that were found to perform less well as well as some minor adaptations to two items to improve their localisation into languages other than English.

The test-retest study described above was conducted using Version 1.0 of DSI. As the two versions correlate highly, these reliability estimates are assumed to hold for Version 1.1 of the instrument.

Evaluating the fairness of DSI scores

The programme that supported the revision to DSI Version 1.1 included a number of analyses at the item level and at the score level to evaluate the fairness of the instrument. We will first describe differential item functioning or DIF procedures used to evaluate whether DSI items operated in an equivalent way across different levels of English language fluency. These procedures were also used to evaluate any potential sources of bias in items by gender, age and ethnicity.

Evaluating differential item functioning (DIF) of DSI items for English fluency

Item level analyses were conducted in South Africa where English is widely used as the language of business but where there are also a number of other languages spoken. As such, a key concern was to identify whether the DSI items would operate equivalently across different levels of fluency in English. Specifically, a series of differential item functioning or DIF analyses were conducted using the procedures described by Zumbo (1999) as well as a number of item p value plots examples of which are provided in Figures 5 and 6. DIF has been defined by Hambleton, Swaminathan and Rogers (1991) as "An item shows DIF if individuals having the same ability, but from different groups, do not have the same probability of getting the item right". DIF analysis serves to evaluate the extent to which items and the scores taken from them place individuals from different groups on the same metric, or whether the unit of measurement upon which people are placed using an instrument is influenced by group membership.

Samples were obtained across three client sites in South Africa covering the automotive, banking and mining industries. The total sample of 381 was used for the item level analyses and these included original DSI Version 1.0 items as well as five adapted items based on conversations with colleagues in South Africa and small focus groups of operational level staff conducted by SHL staff in South Africa. English fluency was categorised as mother tongue (very high), non-mother tongue but very fluent (high), non-mother tongue but fairly fluent (moderate) and non-mother tongue and not very fluent (low). Participants in the study self rated their levels of English fluency. For the DIF analyses, English fluency was recoded into a binary (nominal) variable of high (the first two categories described) versus moderate to low fluency (the last two categories described).

DIF analysis checks were carried out to ensure all psychometric properties of the items were maintained. Of the original 22 items, four were found to perform inconsistently across levels of language fluency. Two amended items were used to replace existing Version 1.0 items as, while item functioning was equivalent, translation checks suggested that the amended items would be easier to localise. For the four items replaced due to inconsistent functioning, all were found to demonstrate moderate levels of uniform DIF and no items were found to exhibit non-uniform DIF (see Zumbo, 1999, for a more detailed explanation of these two types of DIF).

Two examples are shown in Figures 5 and 6 which provide p value plots first for an item showing no DIF and then for an item displaying DIF and that was removed in the process of refining Version 1.1 of DSI. In each figure, the performance of the item is plotted for the two levels of English fluency. The horizontal axis represents total score on the trial form broken down into equal 20% intervals from lowest 20% of scores (1) to the highest 20% of scores (5). The vertical axis represents the probability of people in each overall score interval responding with the preferred answer to the item (i.e. the response option keyed to indicate higher dependability). Please note that item numbering used in Figure 5 and 6 refers to the order in which items were presented in the trial forms used in this study.

Figure 5: P value plot of a DSI item showing equivalent functioning across levels of English fluency

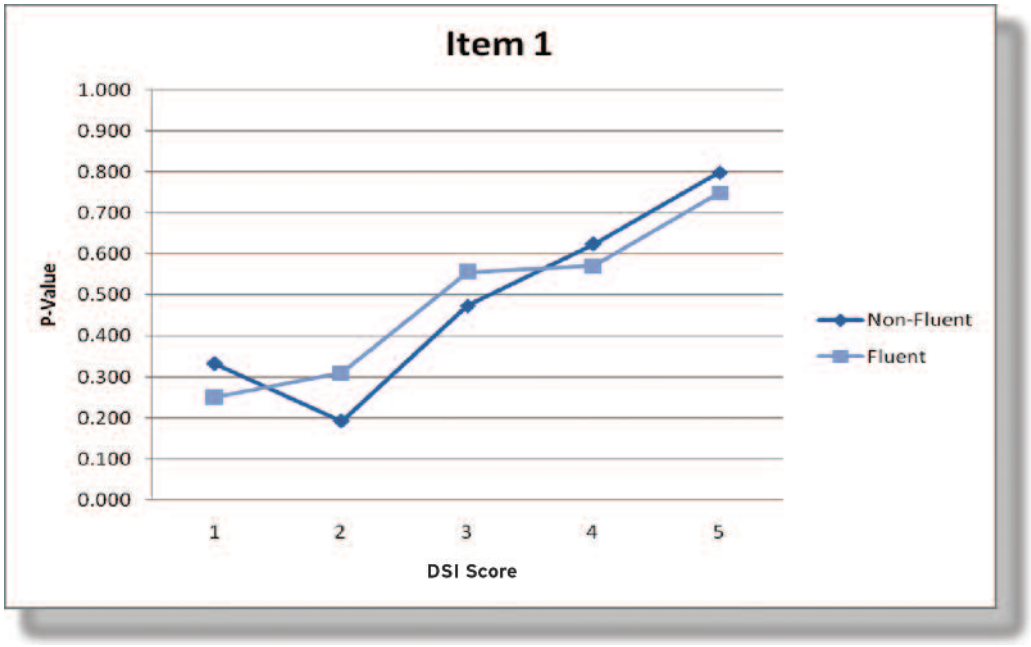
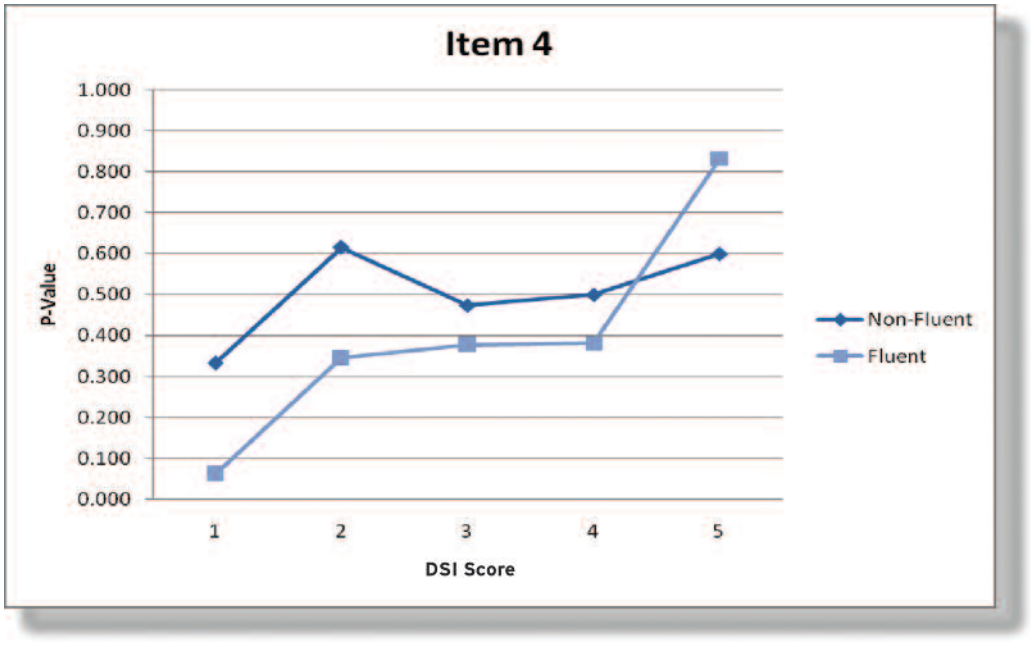


Figure 6: P value plot of a DSI item performing inconsistently across levels of English fluency



We will now describe the DIF procedures in more detail and their application to analyses of item bias by age, gender and ethnicity. We have conducted such analyses in a variety of countries given the local nature of national employment laws, but we will focus on US data for the purposes of exposition in this manual.

Evaluating differential item functioning (DIF) of DSI items for Demographic Groups

We have already described that DIF is concerned with differences in the likelihood of responding to items that is associated with group membership once the construct or trait being measured by an instrument has been taken into account. These analyses are typically applied to explore whether items exhibit bias or DIF associated with demographics of gender or ethnicity, and we extended this concern to age in the analyses conducted for DSI items.

Focusing on data from the US and a total sample size of 430, we will now describe the results of DIF analyses for these demographics. The analyses reported below focused on the final 18 items selected for inclusion in DSI Version 1.1. The results that we will show for US data are consistent with results we have obtained for UK and South African data. Details of the samples used in the DIF analyses are provided below:

Table 17: Data used for DIF analysis (US only)

Demographic	Reference Group	Focal Group	Total Sample For Analysis
Gender	Males = 244 (57%)	Females = 186 (43%)	430
Ethnicity	Whites = 180 (42%)	Non-whites = 243 (58%)	424
Age	Less than 40 years = 265 (62%)	40 years or older = 159 (38%)	423

In the DIF analyses, demographic data were coded 1 for the reference group and 0 for the focal group. So, for the analysis by gender, males were coded 1 and females 0. The analysis for DIF followed the Zumbo approach in which the responses to items are regressed onto three variables; the test score (in this case the overall DSI score), the demographic variable (e.g. males coded 1 and females coded 0) and the interaction between the test score and the demographic variable. For the analyses, DSI items were recoded into 1 if the item pair keyed for dependability was selected and 0 if the other two response options were selected.

The procedure for judging whether DIF is present or not is straightforward as the Zumbo model is a nested model with factors for both uniformed and non-uniformed DIF. The regression analysis provides estimates for all three models where the minimal model is the test score itself (and the results are essentially a form of item-total correlation or discrimination value). The next level of model includes the test score and the demographic variable, and the full or saturated model in this case includes the interaction term in addition to the previous two variables.

Differences in the R^2 values for the first and second models are used to evaluate the presence of uniformed DIF, while differences between the third and second models serve to evaluate the presence of non-uniformed DIF (i.e. that the relationship between item scores and overall scores has a non-additive relationship with the demographic variable).

Zumbo recommends a difference in Multiple R^2 values of 0.13 (equivalent to a difference in Multiple Rs of 0.36) to declare DIF. We have used a more conservative estimate of an R^2 differences of 0.1 to declare DIF (i.e. a difference in R of 0.3). As such, we have applied a more stringent test of DIF to DSI items.

Results for gender.

- Only two items were found to approximate moderate levels of uniformed DIF with R^2 differences of 0.071 and 0.073 respectively for the first and second models in the analysis
- However, while neither item met the criterion set for DIF, each item showed bias in opposite directions (one to males and one to females) effectively cancelling any bias effect out in the overall DSI score
- **As such, the evidence suggests no substantial item bias associated with gender for DSI items.**

Results by ethnicity.

- No items were found to meet the criteria for either uniformed or non-uniformed DIF. We continue to collect data to allow us to conduct more detailed analysis between specific ethnic groups (e.g. Whites and Blacks as they would be defined under US Equal Employment Opportunities Commission guidelines)
- The finding of no item bias by ethnicity permitted by the current data set is consistent with the results of other DIF analyses performed on DSI, and the lack of differences found by ethnicity for personality tests
- **As such, our results suggest no evidence for consistent or substantial item bias associated with ethnicity for DSI items.**

Results by age.

- Age was coded as 1 for those 39 years or less and 0 for those 40 years or more as is consistent with US equality guidelines
- The analyses showed one item approaching uniform DIF with a R^2 difference of 0.083
- **As such, our results suggest no evidence for consistent or substantial item bias associated with age.**

Overall, the results of the US data indicate that DSI items operate equivalently irrespective of whether candidates are male or female, white or non-white, older or younger. As has been mentioned earlier in this section, details of similar analyses will be provided in technical supplements by language or country. In the next section, we describe the extension of our analyses to looking at fairness at the score level and the issue of adverse impact.

Evaluating adverse (disparate) impact of applying DSI risk bands

While DIF analyses serve to evaluate whether the same score metric can be used with different groups, another source of evidence on the fairness of an instrument relates to the use of that instrument to make decisions such as to hire or not to hire an applicant. This is of particular concern in the US which has amongst the most developed principles for the fair use of tests in recruitment and selection, and perhaps the most developed case law in this area (see Landy, 2005, for detailed commentaries and case studies). US litigation on the fairness of a selection process and the assessments used within it tend to hinge around two issues:

- **Disparate treatment** which hinges on whether the candidate was treated differently, whether different treatment can be shown to have been unfair, and whether that treatment was inappropriately related to the candidate's ethnicity or race, religion, sex, age or disability. In a disparate treatment case, the applicant is required to show that the employer's rationale for the employment practice lacks credibility and that the basis for the practice is discriminatory. To respond to such a claim, the employer is required to provide evidence of the logic behind the practice, and that logic needs to be backed up by data that shows the employment practice is not discriminatory.
- **Disparate impact** arises when an employer introduces a practice that, while not intentionally discriminatory, is claimed to exclude or adversely affect members of groups protected under employment law. This is the form of discrimination most closely associated with assessment and is often referred to as adverse impact. In a case of disparate impact, proof of the claim relies on the applicant showing that an alternative and equally valid process would have resulted in lower or no adverse impact. Responses to such claims may require the employer to provide statistical evidence that the process is not systematically biased, which takes us back to making sure that the science is good and the evidence that it is good has been collected.

We will focus on data related to disparate or adverse impact as this is one of the most frequent issues raised in the use of assessments in recruitment and selection. Under US employment law, a rule of thumb that is widely used is the 80% or 4/5th's rule, under which a case for adverse impact may exist if the proportion of a group protected under US employment law, often referred to as the focal group, is less than 80% or 4/5th's of the proportion selected of a majority group, often referred to as the reference group (see Outtz, 2010, for a detailed treatment of adverse impact definitions and issues).

We will explore scenarios for using DSI score bands with four different national data sets; those from Australia, South Africa, the UK and the US. In some cases, such as data from Australia, comparisons by demographic group are limited to gender given the access provided to demographic data by the organisations involved. This also explains why the reader will see differences in the samples sizes for demographics by country in the tables presented below.

We will begin by looking at gender and ethnicity and then return to consider age separately given the reciprocal nature of employment laws associated with age. That is, while employment law promoting fairness by age was largely a response to issues related to older job applicants and job incumbents, ageism against younger members of society has more recently become a strong theme in discussions of appropriate and fair treatment of people at work.

The proportion of applicants shown in the following tables are sample specific and reflect the applicants attracted to particular organisations in particular national labour markets. The percentage expected on average for each cut-score is given in the right hand column of each table headed 'Expected % selected'. In some instances, the reader may see differences between Expected % selected and the sample specific percentages reported in the tables. Any differences observed are related to factors such as the recruitment processes or methods for attracting candidates used by different organisations and conditions in local labour markets.

The tables show the percentage of applicants who would qualify depending on the DSI risk band used for screening. For example, where a table shows 'High', all applicants in the 'Very High' DSI band would be screened out. Where a table shows 'Low', all applicants scoring in the 'Very High' to 'Moderate to Low' bands would be screened out. Policies related to the setting of cut-scores should be developed to reflect local market conditions, legal and best practice requirements, as well as where DSI is placed in a recruitment and selection process and organisational requirements.

Adverse impact and Australian data sets. Two data sets provided data for comparisons by gender, one from the hospitality industry and one from engineering. No data were available from either organisation for ethnicity, but one data set did supply data by age which we will return to later in this section in the discussion of DSI scores and age. The overall (aggregated) sample comprised 80.3% males and 19.7% females. Table 18 shows 4/5th's comparisons by DSI band for this sample. The final column shows whether the selection ratio for females (the focal group) is equal to or greater than 80% of the selection ratio for males (the reference group). The 4/5th's rule is met in all cases.

Table 18: Selection ratios for Australian sample of 122 across two companies broken down by DSI risk band

DSI Risk Band	% males selected	% females selected	Meets 4/5th's rule	Expected % selected
High	99%	96%	Yes	90%
Moderate	87%	83%	Yes	80%
Moderate to Low	68%	63%	Yes	65%
Low	45%	42%	Yes	50%

Note: The lowest DSI risk band (very High Risk) has been excluded from this table as the lowest cut-score that can be used for screening purposes is the High (second) risk band

Adverse impact and South African data sets. Data were available from various client trials as well as DSI usage in screening of in vivo (real) job applications for organisations ranging from mining through banking to the public sector. For gender comparisons, data were available for 1,398 of whom 69.5% were male and 30.5% female. For comparisons by ethnicity, data were available for 175 of whom 30.3% were White, 58.3% were Black and 11.4% were Indian or Coloured as classified by South African ethnic classifications. Table 19 summarises the 4/5th's comparisons for these samples. The figures given in parentheses by ethnicity show selection ratios for Blacks, Indians and Coloureds combined. As shown in Table 19, all comparisons meet the 4/5th's rule.

Table 19: Selection ratios for South African samples of 1,389 (gender) and 175 (ethnicity) across several companies broken down by DSI risk band

DSI Risk Band	% males selected	% females selected	Meets 4/5th's rule	% Whites selected	% Blacks selected	Meets 4/5th's rule	Expected % selected
High	97%	96%	Yes	94%	99% (98%)	Yes (Yes)	90%
Moderate	89%	92%	Yes	77%	96% (93%)	Yes (Yes)	80%
Moderate to Low	66%	71%	Yes	42%	73% (69%)	Yes (Yes)	65%
Low	42%	48%	Yes	17%	47% (44%)	Yes (Yes)	50%

Note: The lowest DSI risk band (very High Risk) has been excluded from this table as the lowest cut-score that can be used for screening purposes is the High (second) risk band

Adverse impact and UK data sets. Table 20 summarises comparisons for UK data gathered from client trials and use of DSI in staff recruitment by gender and ethnicity. Data represent a wide range of organisations including retail, banking, utilities, transportation, engineering and manufacturing, security and emergency services, as well as local public service organisations. Data were available for 3,412 by gender with 66.3% male and 33.7% female, and for 349 by ethnicity with 93.1% White and 6.9% non-White. All comparisons show in Table 20 meet the 4/5th's rule.

Table 20: Selection ratios for UK samples of 3,412 (gender) and 349 (ethnicity) across several organisations broken down by DSI risk band

DSI Risk Band	% males selected	% females selected	Meets 4/5th's rule	% Whites selected	% non-Whites selected	Meets 4/5th's rule	Expected % selected
High	98%	99%	Yes	92%	96%	Yes	90%
Moderate	96%	96%	Yes	79%	83%	Yes	80%
Moderate to Low	86%	88%	Yes	54%	54%	Yes	65%
Low	71%	70%	Yes	31%	29%	Yes	50%

Note: The lowest DSI risk band (very High Risk) has been excluded from this table as the lowest cut-score that can be used for screening purposes is the High (second) risk band

Adverse impact and US data sets. Three US data sets that cover companies operating in rental services (customer agents), construction (drivers and loaders) as well as manufacturing (shop floor operatives were used to evaluate DSI scores against the 4/5th's rule. The gender split in this aggregated sample was 78.8% male and 21.2% female, and 40.4% White, 54.6% Black and 5% Hispanic. Table 21 summarises the results of these analyses which show DSI to operate well against the 4/5th's rule when selection ratios are compared by gender and by major ethnic groupings.

In summary, analyses conducted on South African samples for language fluency and US samples for item bias and adverse impact show promising results for DSI items and overall DSI score bands in supporting fair and equitable recruitment and selection decisions.

Table 21: Selection ratios for US sample of 424 across three companies broken down by DSI risk band

DSI Risk Band	% males selected	% females selected	Meets 4/5th's rule	% Whites selected	% non-Whites selected	Meets 4/5th's rule	Expected % selected
High	96%	97%	Yes	96%	97%	Yes	90%
Moderate	88%	91%	Yes	86%	92%	Yes	80%
Moderate to Low	66%	60%	Yes	63%	63%	Yes	65%
Low	39%	32%	Yes	34%	37%	Yes	50%

Note: The lowest DSI risk band (very High Risk) has been excluded from this table as the lowest cut-score that can be used for screening purposes is the very High risk band

Age and DSI scores

Hattrup and Roberts (2010) show the complexity of definitions of group membership related to adverse impact and to diversity. While they do not delineate age in great detail, their exploration of issues related to classifications and membership of social groupings does highlight the ambiguity in at least some aspects of various nations' employment laws related to fairness in accessing employment opportunities.

This is perhaps more so than in relation to age. While much employment law related to age originated out of concerns for the rights and opportunities of older workers, more recently initiatives and laws related to ageism have emphasised that ageist behaviour may be manifested in negative attitudes and actions against younger people as much as it might be manifested against older people.

We considered this in our review of age and DSI bands. One approach in line with more traditional views of adverse impact and age would be to treat those 40 years of age or more as representing the focal group, and those 39 years of age or more as the reference group. Such an approach is shown in Table 22 which is based on 3,567 cases comprising 2.3% from Australia, 4.9% from South Africa, 88% from the UK and 4.8% from the US. The pattern of results shown in Table 22 typifies results in terms of adverse impact comparisons for all countries when analysed individually.

Table 22: Selection ratios for aggregated sample of 3,567 broken down by age and DSI risk band

DSI Risk Band	% 39 years or less	% 40 years or more	Meets 4/5th's rule	Expected % selected
High	98%	99%	Yes	90%
Moderate	95%	98%	Yes	80%
Moderate to Low	83%	91%	Yes	65%
Low	68%	75%	Yes	50%

Note: The lowest DSI risk band (very High Risk) has been excluded from this table as the lowest cut-score that can be used for screening purposes is the High (second) risk band

Essentially, the trend is that those 40 years of age or more have higher selection ratios than those 39 years of age or less, reflecting a modest but nonetheless positive relationship between DSI score and age ($r=0.13$, $N=3,567$, $p<0.001$, mean age 30 to 34 years with a range from 18 years to 65 years). This would fit the recent finding by Srivastava, John, Oliver, Golsing, and Potter (2003) that suggests that personality continues to develop well beyond the age of 30 years such that higher mean scores on conscientiousness, agreeableness and emotional stability are observed for older age cohorts. This fits with the reciprocal relationship of personality and experience, and could be expected for those with more work experience and thereby more exposure to the basic disciplines and expectations of the world of work.

We explored this by looking at a data set of 749 people for whom DSI scores, age and years of work experience were available. The correlation between age and DSI score approximated that for the larger data set of 3,567 (r=0.19) while the correlations between age and years of work experience was 0.59 (mean age was again 30 to 34 years and mean years of work experience was between 6 and 10 years). With work experience entered first in a step wise regression with DSI as the dependent variable, the Multiple R was 0.191. Age was then entered as the second step and the Multiple R increased to 0.192 which accounted for a 0.038% increase in the variance accounted for, and which was not significant in terms of any increased prediction of DSI scores.

It would seem, therefore, that the relationship between age and DSI scores may well be mediated by a third variable, work experience. This, in turn, would fit a view that differences in DSI scores reflect differences in maturation that would be expected by greater exposure to the expectations of the workplace over time.

A summary of findings on bias and adverse impact analyses of DSI

As has been shown in this section, DSI items and scores have been subject to a number of analyses that show little or no evidence for any bias (DIF) at the item level by demographic, and that show DSI scores generally meet the 4/5th's rule used to evaluate adverse impact. A detailed analysis of the relationship between age and DSI scores suggests that this relationship is mediated by work experience and reflects a greater awareness of what is expected in the workplace which is consistent with recent research on maturation and the development of personality with age.

Faking and DSI

There is a concern in the research literature about the possibility of response bias or faking good in self-report questionnaires, especially when these are used in high stakes scenarios such as screening or selection (hiring) purposes (e.g. Hough and Oswald, 2000). The concern is that some candidates may intentionally distort their scores on self-report measures to ensure a better fit to the job or role in question. In particular, integrity tests have received criticism for their design in that the questions and response formats may encourage candidates to report only positive responses, and the ease of fakability of such instruments (e.g. faking a good profile as described by Sackett and Wanek, 1996).

There are three ways in self-report measures can be designed to tackle the issue of faking good. One approach is through the use of a covert design rather than an overt design of questionnaire. In overt integrity tests, the questions are direct and typically ask respondents about their attitudes towards theft, punctuality or reliability. In contrast, covert integrity tests (or personality-orientated test) tap into personality traits associated with integrity and good job performance (Sackett, Burris & Callaghan, 1989). It is more difficult for respondents to fake on instruments designed in this way. DSI was carefully designed as a covert measure of dependability and reliability for this very reason.

The second approach is through instrument complexity. Multi-scale measures (i.e. questionnaires that measure more than one personality scale) are more complex and therefore more difficult to fake good on. Although DSI reports a single score as per a Criterion Orientated Personality Scale (COPS), it does tap into four criterion scales of dependability as previously mentioned in this manual.

The third approach to minimising faking is through the use of forced-choice response questions rather than normative or Likert scales (Young, White & Heggestad, 2001). In forced-choice formats, statements are carefully worded to have the same level of social desirability and respondents must choose one statement from a selection of statements that is most like them. This design reduces the ability for respondents to distort the image they present and therefore fake a more favourable score. DSI is a forced-choice questionnaire whereby statements were written carefully so that they are equally attractive or socially desirable to candidates.

The recent research literature has focused on candidate attributes and predictors of faking behaviour. Biderman and Nguyen (2004), among others, have found that cognitive ability is related to faking ability on Big 5 personality dimensions. As noted earlier in this manual, the relationship between DSI and cognitive ability was found to be close to zero in a large sample of call centre applicants.

There are two benefits from this finding. First, DSI adds value to an assessment solution that employs cognitive ability tests (ability tests provide data on person-job fit from a "can do" perspective while DSI contributes to person-job fit from a "will do" perspective). Second, the near zero relationship between DSI and cognitive ability supports the fake resistant design of DSI (i.e. candidates with higher scores on cognitive tests and general mental ability obtain scores on DSI comparable to those with lower scores on cognitive ability tests).

Finally, we have compared item-functioning and test-functioning for employees who have completed DSI in a low stakes environment (test trial or internal audit) to candidates who have taken DSI in a high stakes environment (such as job screening and/or selection/hiring into a job). If DSI were subject to high levels of faking-good, then we would expect the items to function differently depending on whether they are used in a low stakes or a high stakes setting, where bias in the items due to faking would be expected in the high stakes setting (i.e. more candidates responding with the higher scored option for any DSI item). Applying the same DIF analyses as described earlier for languages and for demographic group, the results show that DSI items function equivalently irrespective of whether they have been used in a high or a low stakes scenario. This adds further support for the effectiveness of the fake resistant design of DSI.

Faking good is an issue that test developers need to address particularly in an era of increasing usage of online testing. DSI was specifically designed to address this issue and a key part of the development programme as reported in this manual has been to collect and evaluate evidence of the fake resistance that DSI offers. Even though research suggests that as many as 20% of people may fake in the wrong direction (Griffiths & McDaniels, 2006), we hope that the data we have provided here has shown the efforts that have been made to address this issue in the research and development behind DSI.

Using DSI as a human factors audit to provide data on risks in organisations

In the course of this manual, the focus has been primarily on the use of DSI in the recruitment and selection of personnel. We conclude this manual with a final case study that suggests how DSI can be used as an efficient survey tool to gauge levels of behavioural risk in an organisation. While this case study focuses mainly on safety, analogous cases are easy to imagine in organisations where customer facing roles are key to the engagement of the organisation with its clients or stakeholders. As such, whether DSI scores are used at the individual level as in the case study to be described or at a more aggregate level by business unit, location or stage in a business process, we think that DSI offers an easy and effective way for organisations to benchmark and manage levels of behavioural risk.

The case study involves the North British Distillery Company Limited that produces some of the most famous Scotch whisky brands. By the very nature of its manufacturing business, production and warehouse staff are exposed to processes and equipment that, if used or operated incorrectly, can pose risks to health and safety. Employee safety has always been an important issue for the North British Distillery. Indeed, one of the company's core values is a safe working environment. The distillery wanted to increase levels of safety focus within its workforce and identify any risk areas within its operation practices. The distillery implemented a behavioural safety programme into its organisation, which included the use of the DSI.

Within the distillery's production, warehousing and engineering departments, all safety representatives, line managers and team leaders completed the DSI. To quote Glyn Cave, North British Distillery's Employee Development Manager "The results have identified a clear correlation between those employees that scored below average in the test and their safety record to date. Feeding back the results of the DSI allowed us to raise awareness around safety in an objective and consistent way. As a result additional safety training has been given to staff where needed and in some cases operational teams have been 'swapped around' to ensure that they are balanced to minimise risk. Importantly, the production staff are now reporting higher numbers of 'near miss' incidents and learning from these. As a result we have less time lost through accidents."

References

Ackroyd, S., and Thompson, P. (2003). *Organizational misbehaviour*. London: Sage.

Bartram, D., and Brown, A. (2005). *Five factor model (Big Five) OPQ32 report*. Thames Ditton, UK: SHL.

Bartram, D., Brown, A., Fleck, S., Inceoglu, I., and Ward, K. (2006). *OPQ32 Technical Manual*. Thames Ditton, UK: SHL.

Bartram, D., Robertson, I., and Callinan, M. (2002). *Organizational effectiveness: The role of psychology.* Chichester: John Wiley and Sons.

Berry, C. M., Ones, D. S., and Sackett, P. R. (2007). Interpersonnel deviance, organisation deviance and their common correlates: A review and meta-analysis. *Journal of Applied Psychology,* 92, 410-424.

Borman, W. C., and Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10, 99-109.

Broadbent, D. E., Cooper, P. F., Fitzgerald, K. R., and Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British journal of Clinical Psychology*, 21, 1-16.

Burke, E. (2008). Coaching with the OPQ. In J. Passmore (Ed.) Psychometrics in coaching: *Using psychological and psychometric tools for development*. London: Kogan Page.

Burke, E., Fix, C., and Grosvenor, H. (2008). Screening for the Shadow Side of People at Work. Paper presented at the British Psycological Society, Division of Occupational Psycology Conference, Blackpool (UK), January.

Burke, E., and Kirby, L. (2006). *Dependability and safety instrument: Technical manual*. Thames Ditton, UK: SHL

Christiansen, N. D., Burns, G. N., and Montgomery, G. E. (2005). Reconsidering forced-choice formats for applicant personality assessment. *Human Performance*, 18, 267-307.

Confederation of British Industry (2004). *Room for improvement: CBI absence and labour turnover survey*. London: CBI

Confederation of British Industry (2007b). *Consumers will pay a premium for a great reputation*. News release accessed on April 30th. 2009 via www.cbi.org.uk/ndbs/Press.nsf/0363c1f07c6ca12a8025671c00381cc7/1c148e9ea6c3fe4280257394005e0c5d?OpenDocument

Digman, J. M. (1997). Higher-order factors of the Big 5. *Journal of Personality and Social Psychology*, 6, 1246-1256.

Dwyer, J. H. (1983). *Statistical models for the social and behavioural sciences*. New York: Oxford University Press

Future Foundation (2004). *Getting the edge in the new people economy.* London: Future Foundation Ltd.

Gilliland, S.W. & Hale, J. (2005). How do theories of organizational justice inform fair employee selection practices? In J. Greenberg, & J.A. Colquitt (Eds.) *Handbook of organizational justice: Fundamental questions about fairness in the workplace*. Mahwah, NJ: Erlbaum.

Goodman, J. (1999). Quantifying the Impact of Great Customer Service on Profitability. In R. Zernke and J. Woods (Eds.). *Best practices in customer services*. New York: American Management Association.

Gruys, M. L. (1999). *The dimensionality of deviant employee performance in the workplace*. Unpublished doctoral dissertation. Minneapolis, MN: University of Minnesota.

Gruys, M. L., and Sackett, P. R. (2002). Investigating the dimensionality of counter-productive work behaviour. *International Journal of Selection and Assessment*, 11, 30-42.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Hattrup, K., and Roberts, B. G. (2010). What are the criteria for adverse impact? In Outtz, J. L. (Ed). *Adverse impact: Implications for organizational staffing and high stakes selection.* New York: Routledge.

Health & Safety Executive (HSE) (2004). *HSE updates costs to Britain of workplace accidents and work-related ill health*. HSE Press Release: E139:04

Hogan, J., Hogan, R., and Busch, C. M. (1984). How to measure service orientation. *Journal of Applied Psychology*, 69, 167-173

Hollinger, R. C., and Clark, J. P. (1983). *Theft by employees*. Lexington, MA: DC Health and Company: Lexington Books.

Hunter, J. E., and Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications.

Hunter, J. E., and Scmidt, F. L. (1999). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

Jackson, D. N., and Wroblewski, V. R. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13, 371-388.

Judge, T. A., and Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87, 797-807.

Judge, T. A., Bono, J. E., Ilies, R., and Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.

Kolen, M. J., and Brennan, R. L. (2004). T*est equating, scaling and linking: Methods and practices (2nd. Edition)*. New York: Springer

Landy, F. J. (2005). *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives*. San Francisco: Jossey Bass

LePine, J. A., Erez, A., and Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behaviour: a critical review and meta-analysis. *Journal of Applied Psychology*, 87, 52-65.

Marcus, B., Lee, K., and Ashton, M. C. (2007). Personality dimensions explaining the relationship between integrity tests and counter-productive behaviour: Big five or one in addition? *Personnel Psychology*, 60, 1-34.

Ones, D. S. (1993). *The construct of integrity tests*. Unpublished doctoral dissertation. Iowa City, Iowa: University of Iowa.

Ones, D. S., Viswersveran, C. (2001a). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31-39.

Ones, D. S., and Viswesvaran, C. (2001b). Personality at work: Criterion-focused occupational psychology scales used in personnel selection. In B. W. Roberts and R. Hogan (Eds.). *Personality in the workplace*. Washington D.C.: American Psychological Association.

Outtz, J. L. (2010). *Adverse impact: Implications for organizational staffing and high stakes selection*. New York: Routledge.

Robinson, S. L., and Bennett, R. J. (1995). A typology of deviant workplace behaviours: A multidimensional scaling study. *Academy of Management Journal*, 38, 555-572.

Sackett, P. R. (2002). The structure of counter-productive work behaviours: Dimensionality and relationships with facets of job performance. *International Journal of Selection and Assessment*, 10, 5-11.

Sackett, P. R., and Devore, C. J. (2001). Counter-productive behaviours at work. In N. Anderson, D. S. Ones, H. Sinangil Kepir and C. Viswesvaran (Eds.). *Handbook of Industrial Work and Organisational Psychology: Voumel 1. Personnel Psychology*. London: Sage.

Sackett, P. R., and Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness and reliability for personnel selection. *Personnel Psychology*, 49, 787-829.

Salgado, J. F. (2002). The big five personality dimensions and counter-productive behaviours. *International Journal of Selection and Assessment*, 10, 117-123.

Srivastava, S., John, Oliver P., Golsing, S. D., and Potter, J. (2003). Development of Personality in Early and Middle Adulthood: Set Like Plaster or Persistent Change? *Journal of Personality and Social Psychology,* 84, 1041–1053.

Slora, K. B. (1991). An empirical approach to determining employee deviance base rates. *Journal of Business and Psychology*, 4, 199-219.

SHL (1999). *Work Styles Questionnaire: Manual and user's guide*. Thames Ditton: SHL Group Limited.

Taylor, P. J., Pajo, K., Cheung, G. W., and Stringfield, P. (2005). Dimensionality and validity of a structured reference check procedure. *Personnel Psychology*, 57, 745-772.

Viswesvaran, C. (2002). Absenteeism and measures of job performance: A meta-analysis. *International Journal of Selection and Assessment*, 10, 12-16.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

© 2010, SHL Group Limited
www.shl.com

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means or stored in a database or retrieval system without the prior written permission of SHL Group Limited.

6025